9 | MOLECULAR BIOLOGY

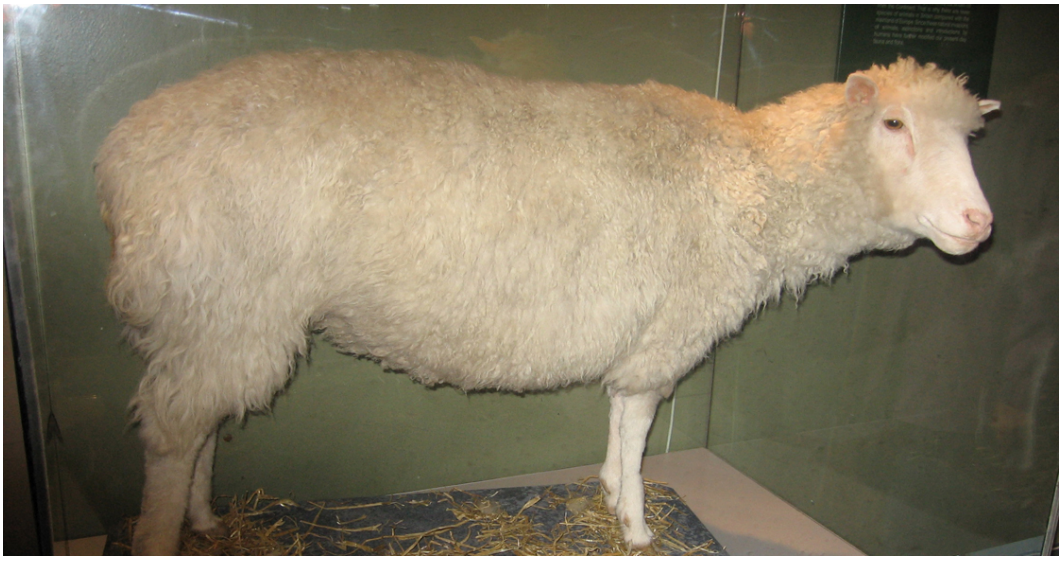


Figure 9.1 Dolly the sheep was the first cloned mammal.

Chapter Outline

Introduction

The three letters “DNA” have now become associated with crime solving, paternity testing, human identification, and genetic testing. DNA can be retrieved from hair, blood, or saliva. With the exception of identical twins, each person’s DNA is unique and it is possible to detect differences between human beings on the basis of their unique DNA sequence.

DNA analysis has many practical applications beyond forensics and paternity testing. DNA testing is used for tracing genealogy and identifying pathogens. In the medical field, DNA is used in diagnostics, new vaccine development, and cancer therapy. It is now possible to determine predisposition to many diseases by analyzing genes.

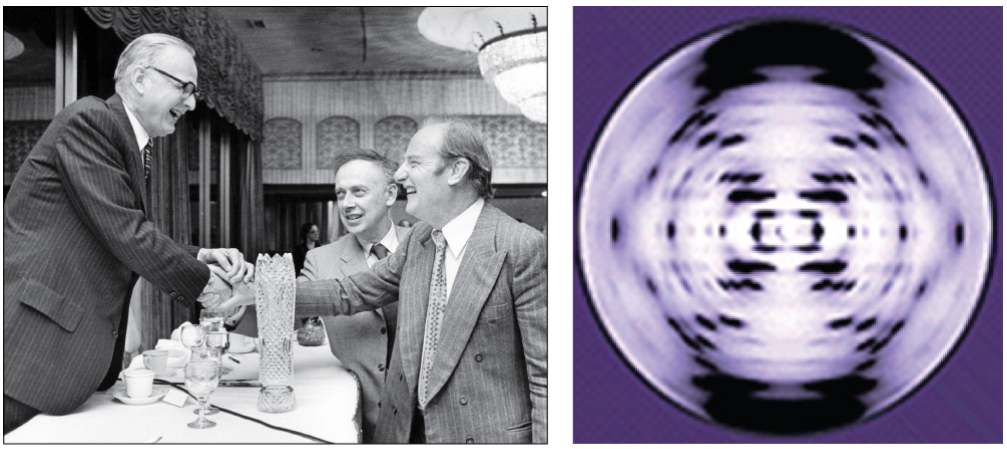
DNA is the genetic material passed from parent to offspring for all life on Earth. The technology of molecular genetics developed in the last half century has enabled us to see deep into the history of life to deduce the relationships between living things in ways never thought possible. It also allows us to understand the workings of evolution in populations of organisms. Over a thousand species have had their entire genome sequenced, and there have been thousands of individual human genome sequences completed. These sequences will allow us to understand human disease and the relationship of humans to the rest of the tree of life. Finally, molecular genetics techniques have revolutionized plant and animal breeding for human agricultural needs. All of these advances in biotechnology depended on basic research leading to the discovery of the structure of DNA in 1953, and the research since then that has uncovered the details of DNA replication and the complex process leading to the expression of DNA in the form of proteins in the cell.

9.1 | The Structure of DNA

By the end of this section, you will be able to:

- Describe the structure of DNA
- Describe how eukaryotic and prokaryotic DNA is arranged in the cell

In the 1950s, Francis Crick and James Watson worked together at the University of Cambridge, England, to determine the structure of DNA. Other scientists, such as Linus Pauling and Maurice Wilkins, were also actively exploring this field. Pauling had discovered the secondary structure of proteins using X-ray crystallography. X-ray crystallography is a method for investigating molecular structure by observing the patterns formed by X-rays shot through a crystal of the substance. The patterns give important information about the structure of the molecule of interest. In Wilkins' lab, researcher Rosalind Franklin was using X-ray crystallography to understand the structure of DNA. Watson and Crick were able to piece together the puzzle of the DNA molecule using Franklin's data (**Figure 9.2**). Watson and Crick also had key pieces of information available from other researchers such as Chargaff's rules. Chargaff had shown that of the four kinds of monomers (nucleotides) present in a DNA molecule, two types were always present in equal amounts and the remaining two types were also always present in equal amounts. This meant they were always paired in some way. In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine for their work in determining the structure of DNA.



(a) (b)

Figure 9.2 Pioneering scientists (a) James Watson and Francis Crick are pictured here with American geneticist Maclyn McCarty. Scientist Rosalind Franklin discovered (b) the X-ray diffraction pattern of DNA, which helped to elucidate its double helix structure. (credit a: modification of work by Marjorie McCarty; b: modification of work by NIH)

Now let's consider the structure of the two types of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The building blocks of DNA are nucleotides, which are made up of three parts: a **deoxyribose** (5-carbon sugar), a **phosphate group**, and a **nitrogenous base** (**Figure 9.3**). There are four types of nitrogenous bases in DNA. Adenine (A) and guanine (G) are double-ringed purines, and cytosine (C) and thymine (T) are smaller, single-ringed pyrimidines. The nucleotide is named according to the nitrogenous base it contains.

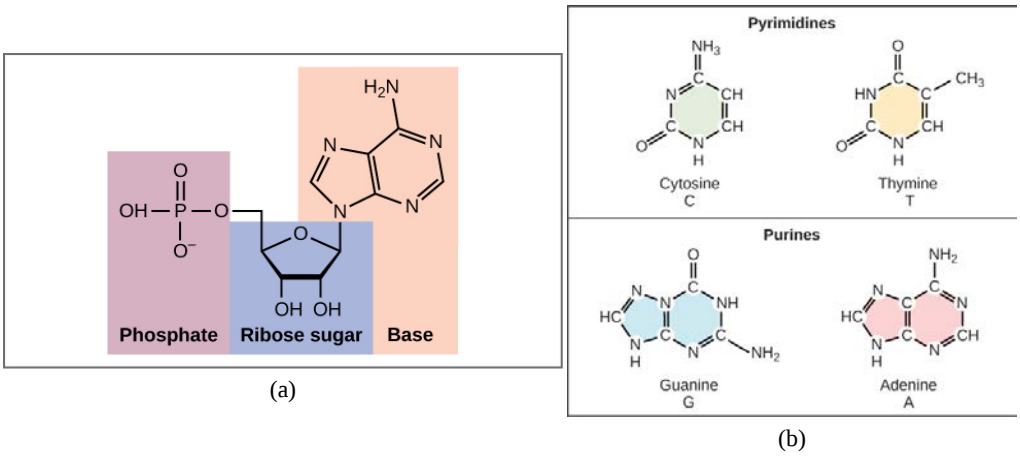


Figure 9.3 (a) Each DNA nucleotide is made up of a sugar, a phosphate group, and a base. (b) Cytosine and thymine are pyrimidines. Guanine and adenine are purines.

The phosphate group of one nucleotide bonds covalently with the sugar molecule of the next nucleotide, and so on, forming a long polymer of nucleotide monomers. The sugar–phosphate groups line up in a "backbone" for each single strand of DNA, and the nucleotide bases stick out from this backbone. The carbon atoms of the five-carbon sugar are numbered clockwise from the oxygen as 1', 2', 3', 4', and 5' (1' is read as "one prime"). The phosphate group is attached to the 5' carbon of one nucleotide and the 3' carbon of the next nucleotide. In its natural state, each DNA molecule is actually composed of two single strands held together along their length with hydrogen bonds between the bases.

Watson and Crick proposed that the DNA is made up of two strands that are twisted around each other to form a right-handed helix, called a **double helix**. Base-pairing takes place between a purine and pyrimidine: namely, A pairs with T, and G pairs with C. In other words, adenine and thymine are complementary base pairs, and cytosine and guanine are also complementary base pairs. This is the basis for Chargaff's rule; because of their complementarity, there is as much adenine as thymine in a DNA molecule and as much guanine as cytosine. Adenine and thymine are connected by two hydrogen bonds, and cytosine and guanine are connected by three hydrogen bonds. The two strands are anti-parallel in nature; that is, one strand will have the 3' carbon of the sugar in the "upward" position, whereas the other strand will have the 5' carbon in the upward position. The diameter of the DNA double helix is uniform throughout because a purine (two rings) always pairs with a pyrimidine (one ring) and their combined lengths are always equal. (**Figure 9.4**).

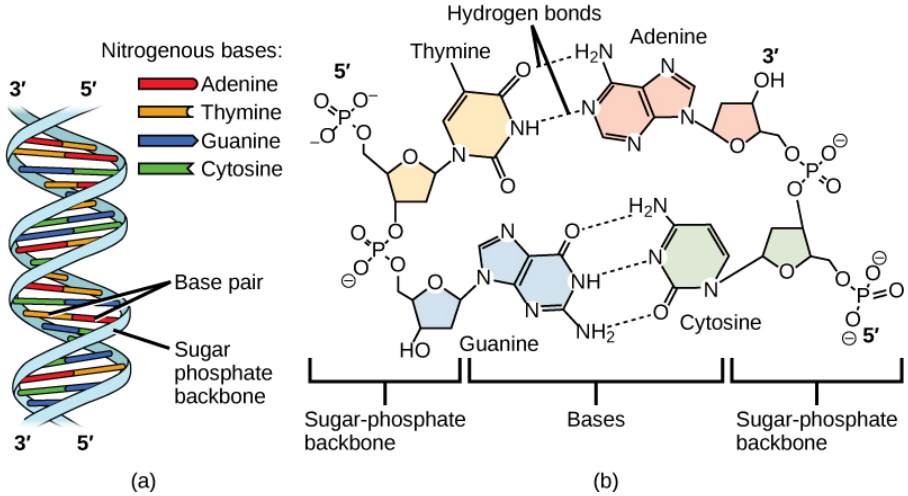


Figure 9.4 DNA (a) forms a double stranded helix, and (b) adenine pairs with thymine and cytosine pairs with guanine. (credit a: modification of work by Jerome Walker, Dennis Myts)

The Structure of RNA

There is a second nucleic acid in all cells called ribonucleic acid, or RNA. Like DNA, RNA is a polymer of nucleotides. Each of the nucleotides in RNA is made up of a nitrogenous base, a five-carbon sugar, and a phosphate group. In the case of RNA, the five-carbon sugar is ribose, not deoxyribose. Ribose has a hydroxyl group at the 2' carbon, unlike deoxyribose, which has only a hydrogen atom (**Figure 9.5**).

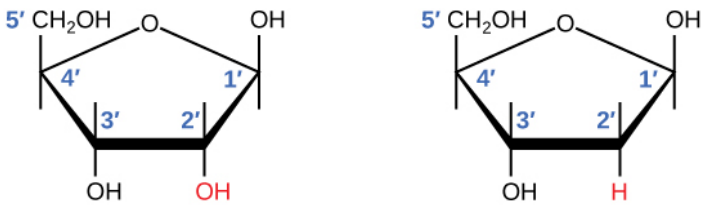


Figure 9.5 The difference between the ribose found in RNA and the deoxyribose found in DNA is that ribose has a hydroxyl group at the 2' carbon.

RNA nucleotides contain the nitrogenous bases adenine, cytosine, and guanine. However, they do not contain thymine, which is instead replaced by uracil, symbolized by a "U." RNA exists as a single-stranded molecule rather than a double-stranded helix. Molecular biologists have named several kinds of RNA on the basis of their function. These include messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA)—molecules that are involved in the production of proteins from the DNA code.

How DNA Is Arranged in the Cell

DNA is a working molecule; it must be replicated when a cell is ready to divide, and it must be "read" to produce the molecules, such as proteins, to carry out the functions of the cell. For this reason, the DNA is protected and packaged in very specific ways. In addition, DNA molecules can be very long. Stretched end-to-end, the DNA molecules in a single human cell would come to a length of about 2 meters. Thus, the DNA for a cell must be packaged in a very ordered way to fit and function within a structure (the cell) that is not visible to the naked eye. The chromosomes of prokaryotes are much simpler than those of eukaryotes in many of their features (**Figure 9.6**). Most prokaryotes contain a single, circular chromosome that is found in an area in the cytoplasm called the nucleoid.

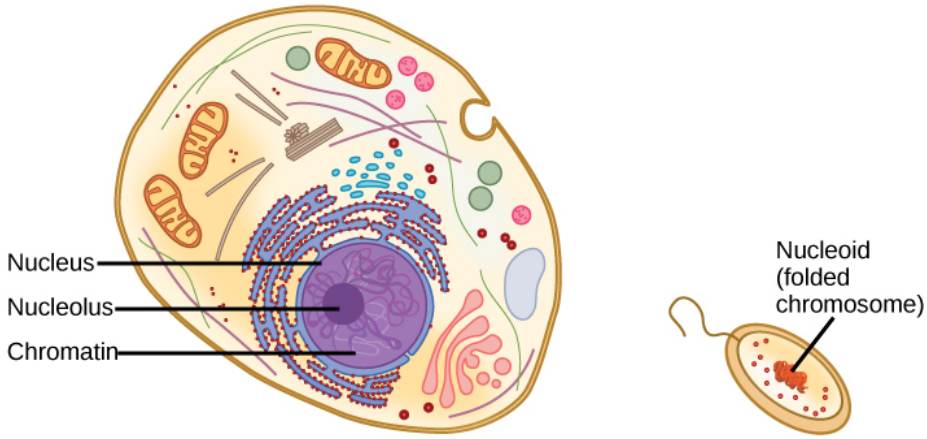


Figure 9.6 A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

The size of the genome in one of the most well-studied prokaryotes, *Escherichia coli*, is 4.6 million base pairs, which would extend a distance of about 1.6 mm if stretched out. So how does this fit inside a small bacterial cell? The DNA is twisted beyond the double helix in what is known as supercoiling. Some proteins are known to be involved in the supercoiling; other proteins and enzymes help in maintaining the supercoiled structure.

Eukaryotes, whose chromosomes each consist of a linear DNA molecule, employ a different type of packing strategy to fit their DNA inside the nucleus (**Figure 9.7**). At the most basic level, DNA is wrapped around proteins known as histones to form structures called nucleosomes. The DNA is wrapped tightly around the histone core. This nucleosome is linked to the next one by a short strand of DNA that is free of histones. This is also known as the "beads on a string" structure; the nucleosomes are the "beads" and the short lengths of DNA between them are the "string." The nucleosomes, with their DNA coiled around them, stack compactly onto each other to form a 30-nm–wide fiber. This fiber is further coiled into a thicker and more compact structure. At the metaphase stage of mitosis, when the chromosomes are lined up in the center of the cell, the chromosomes are at their most compacted. They are approximately 700 nm in width, and are found in association with scaffold proteins.

In interphase, the phase of the cell cycle between mitoses at which the chromosomes are decondensed, eukaryotic chromosomes have two distinct regions that can be distinguished by staining. There is a tightly packaged region that stains darkly, and a less dense region. The darkly staining regions usually contain genes that are not active, and are found in the regions of the centromere and telomeres. The lightly staining regions usually contain genes that are active, with DNA packaged around nucleosomes but not further compacted.

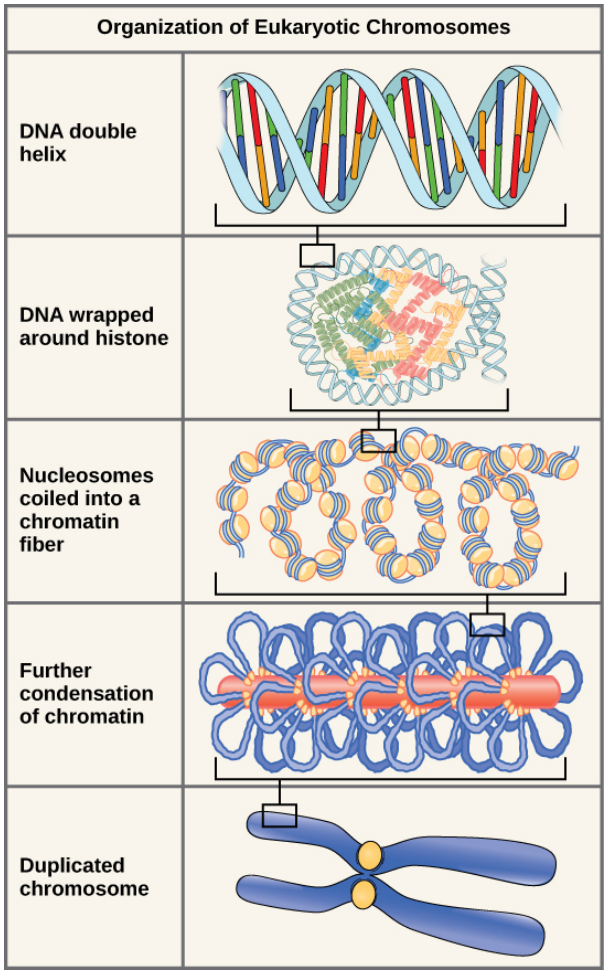


Figure 9.7 These figures illustrate the compaction of the eukaryotic chromosome.

CONCEPT in ACTION

Watch this **animation (http://openstaxcollege.org/l/DNA_packaging)** of DNA packaging.

9.2 | DNA Replication

By the end of this section, you will be able to:

- Explain the process of DNA replication
- Explain the importance of telomerase to DNA replication
- Describe mechanisms of DNA repair

When a cell divides, it is important that each daughter cell receives an identical copy of the DNA. This is accomplished by the process of DNA replication. The replication of DNA occurs during the synthesis phase, or S phase, of the cell cycle, before the cell enters mitosis or meiosis.

The elucidation of the structure of the double helix provided a hint as to how DNA is copied. Recall that adenine nucleotides pair with thymine nucleotides, and cytosine with guanine. This means that the two strands are complementary to each other. For example, a strand of DNA with a nucleotide sequence of AGTCATGA will have a complementary strand with the sequence TCAGTACT (**Figure 9.8**).

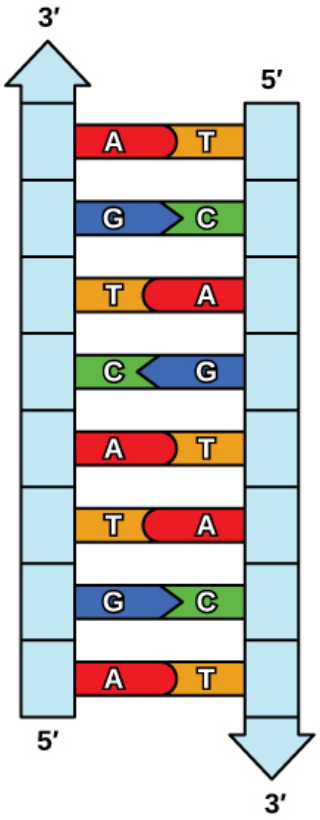


Figure 9.8 The two strands of DNA are complementary, meaning the sequence of bases in one strand can be used to create the correct sequence of bases in the other strand.

Because of the complementarity of the two strands, having one strand means that it is possible to recreate the other strand. This model for replication suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied (**Figure 9.9**).

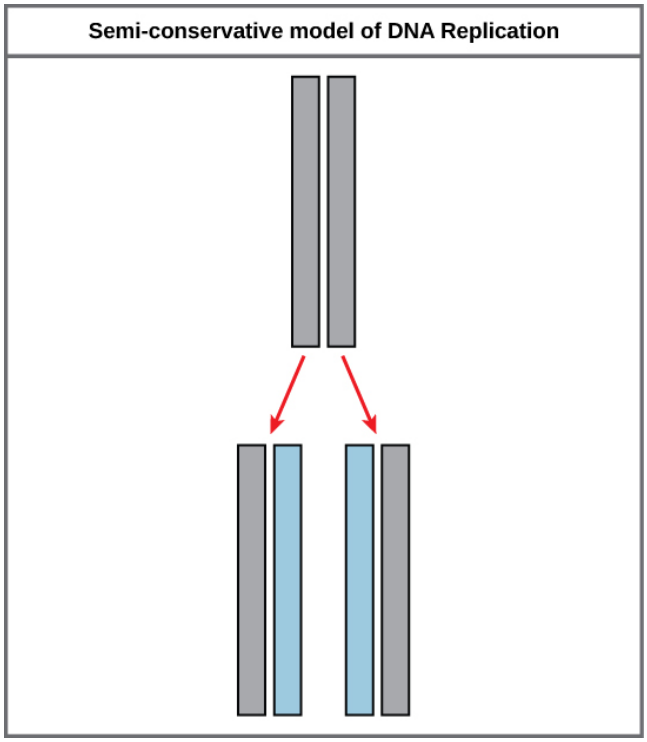


Figure 9.9 The semiconservative model of DNA replication is shown. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

During DNA replication, each of the two strands that make up the double helix serves as a template from which new strands are copied. The new strand will be complementary to the parental or “old” strand. Each new double strand consists of one parental strand and one new daughter strand. This is known as **semiconservative replication**. When two DNA copies are formed, they have an identical sequence of nucleotide bases and are divided equally into two daughter cells.

DNA Replication in Eukaryotes

Because eukaryotic genomes are very complex, DNA replication is a very complicated process that involves several enzymes and other proteins. It occurs in three main stages: initiation, elongation, and termination.

Recall that eukaryotic DNA is bound to proteins known as histones to form structures called nucleosomes. During initiation, the DNA is made accessible to the proteins and enzymes involved in the replication process. How does the replication machinery know where on the DNA double helix to begin? It turns out that there are specific nucleotide sequences called origins of replication at which replication begins. Certain proteins bind to the origin of replication while an enzyme called **helicase** unwinds and opens up the DNA helix. As the DNA opens up, Y-shaped structures called **replication forks** are formed (**Figure 9.10**). Two replication forks are formed at the origin of replication, and these get extended in both directions as replication proceeds. There are multiple origins of replication on the eukaryotic chromosome, such that replication can occur simultaneously from several places in the genome.

During elongation, an enzyme called **DNA polymerase** adds DNA nucleotides to the 3' end of the template. Because DNA polymerase can only add new nucleotides at the end of a backbone, a **primer** sequence, which provides this starting point, is added with complementary RNA nucleotides. This primer is removed later, and the nucleotides are replaced with DNA nucleotides. One strand, which is complementary to the parental DNA strand, is synthesized continuously toward the replication fork so the polymerase can add nucleotides in this direction. This continuously synthesized strand is known as the **leading strand**. Because DNA polymerase can only synthesize DNA in a 5' to 3' direction, the other new strand is put together in short pieces called **Okazaki fragments**. The Okazaki fragments each require a primer made of RNA to start the synthesis. The strand with the Okazaki fragments is known as the **lagging strand**. As synthesis proceeds, an enzyme removes the RNA primer, which is then replaced with DNA nucleotides, and the gaps between fragments are sealed by an enzyme called **DNA ligase**.

The process of DNA replication can be summarized as follows:

1. DNA unwinds at the origin of replication.

2. New bases are added to the complementary parental strands. One new strand is made continuously, while the other strand is made in pieces.
3. Primers are removed, new DNA nucleotides are put in place of the primers and the backbone is sealed by DNA ligase.

art CONNECTION

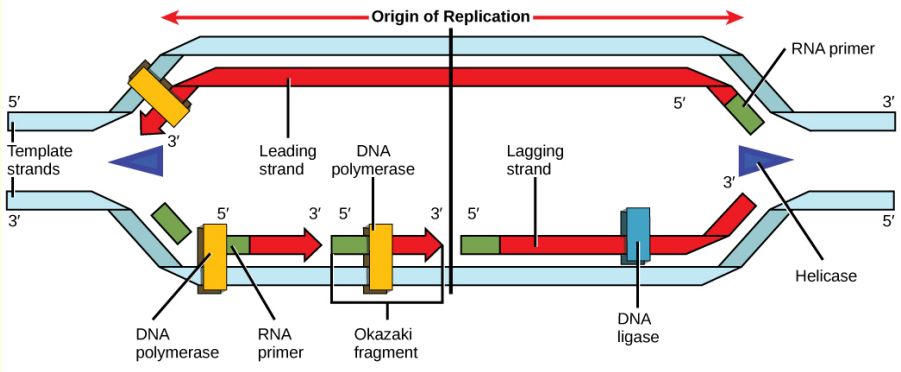


Figure 9.10 A replication fork is formed by the opening of the origin of replication, and helicase separates the DNA strands. An RNA primer is synthesized, and is elongated by the DNA polymerase. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches. The DNA fragments are joined by DNA ligase (not shown).

You isolate a cell strain in which the joining together of Okazaki fragments is impaired and suspect that a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

Telomere Replication

Because eukaryotic chromosomes are linear, DNA replication comes to the end of a line in eukaryotic chromosomes. As you have learned, the DNA polymerase enzyme can add nucleotides in only one direction. In the leading strand, synthesis continues until the end of the chromosome is reached; however, on the lagging strand there is no place for a primer to be made for the DNA fragment to be copied at the end of the chromosome. This presents a problem for the cell because the ends remain unpaired, and over time these ends get progressively shorter as cells continue to divide. The ends of the linear chromosomes are known as **telomeres**, which have repetitive sequences that do not code for a particular gene. As a consequence, it is telomeres that are shortened with each round of DNA replication instead of genes. For example, in humans, a six base-pair sequence, TTAGGG, is repeated 100 to 1000 times. The discovery of the enzyme **telomerase** (**Figure 9.11**) helped in the understanding of how chromosome ends are maintained. The telomerase attaches to the end of the chromosome, and complementary bases to the RNA template are added on the end of the DNA strand. Once the lagging strand template is sufficiently elongated, DNA polymerase can now add nucleotides that are complementary to the ends of the chromosomes. Thus, the ends of the chromosomes are replicated.

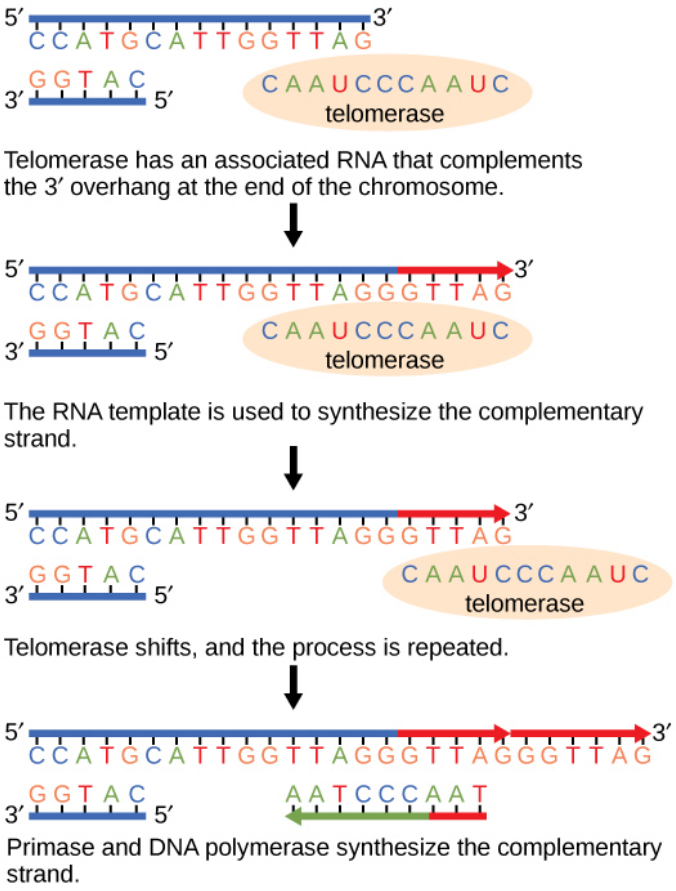


Figure 9.11 The ends of linear chromosomes are maintained by the action of the telomerase enzyme.

Telomerase is typically found to be active in germ cells, adult stem cells, and some cancer cells. For her discovery of telomerase and its action, Elizabeth Blackburn (**Figure 9.12**) received the Nobel Prize for Medicine and Physiology in 2009.

Figure 9.12 Elizabeth Blackburn, 2009 Nobel Laureate, was the scientist who discovered how telomerase works. (credit: U.S. Embassy, Stockholm, Sweden)

Telomerase is not active in adult somatic cells. Adult somatic cells that undergo cell division continue to have their telomeres shortened. This essentially means that telomere shortening is associated with aging. In 2010, scientists found that telomerase can reverse some age-related conditions in mice, and this may have potential in regenerative medicine.[1]

1. Mariella Jaskelioff, et al., "Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice," *Nature*, 469 (2011):102–7.

Telomerase-deficient mice were used in these studies; these mice have tissue atrophy, stem-cell depletion, organ system failure, and impaired tissue injury responses. Telomerase reactivation in these mice caused extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved functioning of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

DNA Replication in Prokaryotes

Recall that the prokaryotic chromosome is a circular molecule with a less extensive coiling structure than eukaryotic chromosomes. The eukaryotic chromosome is linear and highly coiled around proteins. While there are many similarities in the DNA replication process, these structural differences necessitate some differences in the DNA replication process in these two life forms.

DNA replication has been extremely well-studied in prokaryotes, primarily because of the small size of the genome and large number of variants available. *Escherichia coli* has 4.6 million base pairs in a single circular chromosome, and all of it gets replicated in approximately 42 minutes, starting from a single origin of replication and proceeding around the chromosome in both directions. This means that approximately 1000 nucleotides are added per second. The process is much more rapid than in eukaryotes. **Table 9.1** summarizes the differences between prokaryotic and eukaryotic replications.

Differences between Prokaryotic and Eukaryotic Replications

Property	Prokaryotes	Eukaryotes
Origin of replication	Single	Multiple
Rate of replication	1000 nucleotides/s	50 to 100 nucleotides/s
Chromosome structure	circular	linear
Telomerase	Not present	Present

Table 9.1

Click through a **tutorial (http://openstaxcollege.org/l/DNA_replicatio2)** on DNA replication.

DNA Repair

DNA polymerase can make mistakes while adding nucleotides. It edits the DNA by proofreading every newly added base. Incorrect bases are removed and replaced by the correct base, and then polymerization continues (**Figure 9.13a**). Most mistakes are corrected during replication, although when this does not happen, the **mismatch repair** mechanism is employed. Mismatch repair enzymes recognize the wrongly incorporated base and excise it from the DNA, replacing it with the correct base (**Figure 9.13b**). In yet another type of repair, **nucleotide excision repair**, the DNA double strand is unwound and separated, the incorrect bases are removed along with a few bases on the 5' and 3' end, and these are replaced by copying the template with the help of DNA polymerase (**Figure 9.13c**). Nucleotide excision repair is particularly important in correcting thymine dimers, which are primarily caused by ultraviolet light. In a thymine dimer, two thymine nucleotides adjacent to each other on one strand are covalently bonded to each other rather than their complementary bases. If the dimer is not removed and repaired it will lead to a mutation. Individuals with flaws in their nucleotide excision repair genes show extreme sensitivity to sunlight and develop skin cancers early in life.

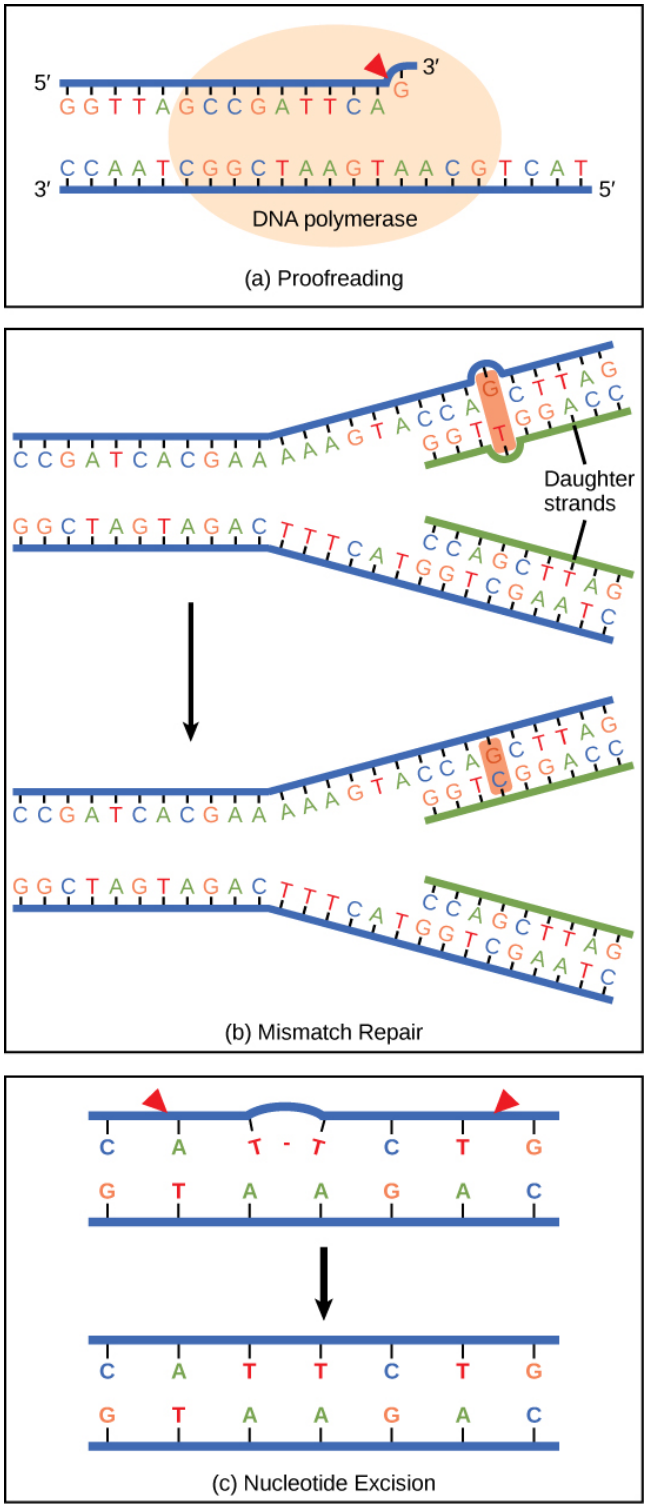


Figure 9.13 Proofreading by DNA polymerase (a) corrects errors during replication. In mismatch repair (b), the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base. Nucleotide excision (c) repairs thymine dimers. When exposed to UV, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced.

Most mistakes are corrected; if they are not, they may result in a **mutation**—defined as a permanent change in the DNA sequence. Mutations in repair genes may lead to serious consequences like cancer.

9.3 | Transcription

By the end of this section, you will be able to:

- Explain the central dogma
- Explain the main steps of transcription
- Describe how eukaryotic mRNA is processed

In both prokaryotes and eukaryotes, the second function of DNA (the first was replication) is to provide the information needed to construct the proteins necessary so that the cell can perform all of its functions. To do this, the DNA is "read" or transcribed into an **mRNA** molecule. The mRNA then provides the code to form a protein by a process called translation. Through the processes of transcription and translation, a protein is built with a specific sequence of amino acids that was originally encoded in the DNA. This module discusses the details of transcription.

The Central Dogma: DNA Encodes RNA; RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the central dogma (Figure 9.14), which states that genes specify the sequences of mRNAs, which in turn specify the sequences of proteins.

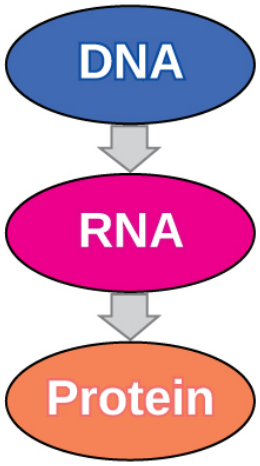


Figure 9.14 The central dogma states that DNA encodes RNA, which in turn encodes protein.

The copying of DNA to mRNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every complementary nucleotide read in the DNA strand. The translation to protein is more complex because groups of three mRNA nucleotides correspond to one amino acid of the protein sequence. However, as we shall see in the next module, the translation to protein is still systematic, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.

Transcription: from DNA to mRNA

Both prokaryotes and eukaryotes perform fundamentally the same process of transcription, with the important difference of the membrane-bound nucleus in eukaryotes. With the genes bound in the nucleus, transcription occurs in the nucleus of the cell and the mRNA transcript must be transported to the cytoplasm. The prokaryotes, which include bacteria and archaea, lack membrane-bound nuclei and other organelles, and transcription occurs in the cytoplasm of the cell. In both prokaryotes and eukaryotes, transcription occurs in three main stages: initiation, elongation, and termination.

Initiation

Transcription requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a **transcription bubble**. The DNA sequence onto which the proteins and enzymes involved in transcription bind to initiate the process is called a **promoter**. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very important because it determines whether the corresponding gene is transcribed all of the time, some of the time, or hardly at all (Figure 9.15).

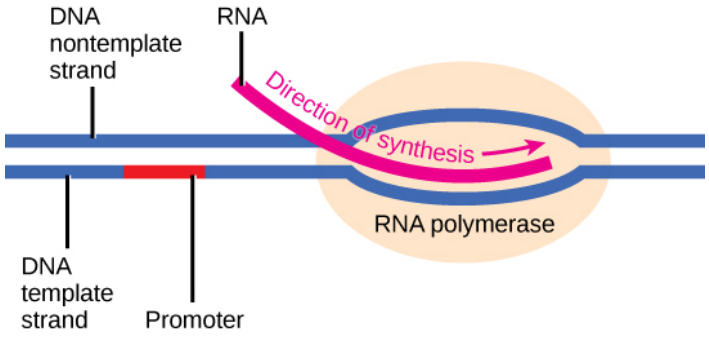


Figure 9.15 The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

Elongation

Transcription always proceeds from one of the two DNA strands, which is called the **template strand**. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the **nontemplate strand**, with the exception that RNA contains a uracil (U) in place of the thymine (T) found in DNA. During elongation, an enzyme called **RNA polymerase** proceeds along the DNA template adding nucleotides by base pairing with the DNA template in a manner similar to DNA replication, with the difference that an RNA strand is being synthesized that does not remain bound to the DNA template. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it (**Figure 9.16**).

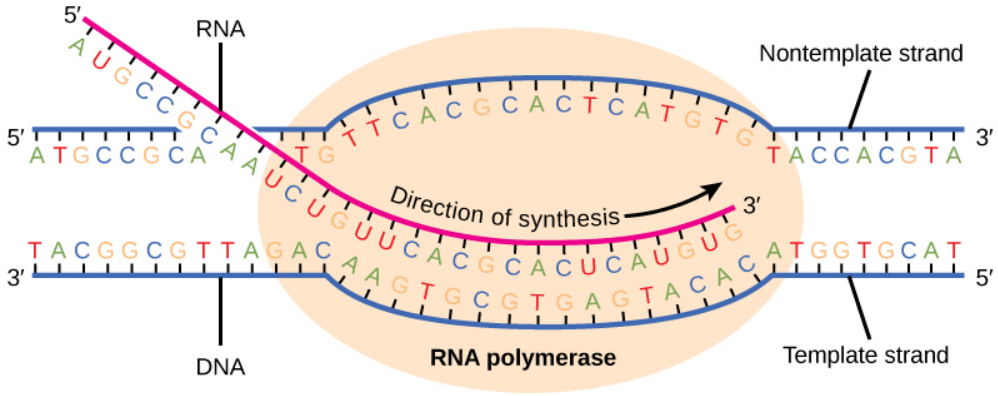


Figure 9.16 During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds then rewinds the DNA as it is read.

Termination

Once a gene is transcribed, the prokaryotic polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals, but both involve repeated nucleotide sequences in the DNA template that result in RNA polymerase stalling, leaving the DNA template, and freeing the mRNA transcript.

On termination, the process of transcription is complete. In a prokaryotic cell, by the time termination occurs, the transcript would already have been used to partially synthesize numerous copies of the encoded protein because these processes can occur concurrently using multiple ribosomes (polyribosomes) (**Figure 9.17**). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.

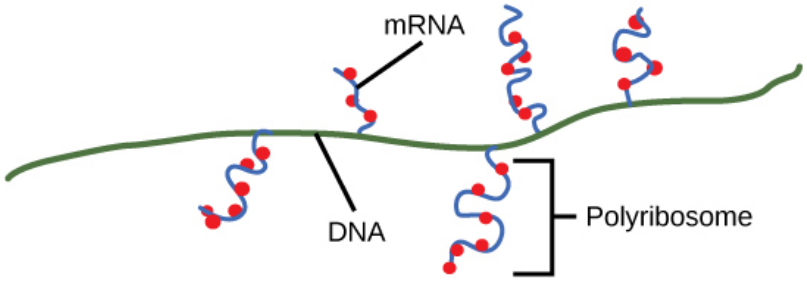


Figure 9.17 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

Eukaryotic RNA Processing

The newly transcribed eukaryotic mRNAs must undergo several processing steps before they can be transferred from the nucleus to the cytoplasm and translated into a protein. The additional steps involved in eukaryotic mRNA maturation create a molecule that is much more stable than a prokaryotic mRNA. For example, eukaryotic mRNAs last for several hours, whereas the typical prokaryotic mRNA lasts no more than five seconds.

The mRNA transcript is first coated in RNA-stabilizing proteins to prevent it from degrading while it is processed and exported out of the nucleus. This occurs while the pre-mRNA still is being synthesized by adding a special nucleotide "cap" to the 5' end of the growing transcript. In addition to preventing degradation, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

Once elongation is complete, an enzyme then adds a string of approximately 200 adenine residues to the 3' end, called the poly-A tail. This modification further protects the pre-mRNA from degradation and signals to cellular factors that the transcript needs to be exported to the cytoplasm.

Eukaryotic genes are composed of protein-coding sequences called **exons** (*ex*-on signifies that they are *ex*pressed) and *int*ervening sequences called **introns** (*int*-ron denotes their *int*ervening role). Introns are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins. It is essential that all of a pre-mRNA's introns be completely and precisely removed before protein synthesis so that the exons join together to code for the correct amino acids. If the process errs by even a single nucleotide, the sequence of the rejoined exons would be shifted, and the resulting protein would be nonfunctional. The process of removing introns and reconnecting exons is called **splicing** (**Figure 9.18**). Introns are removed and degraded while the pre-mRNA is still in the nucleus.

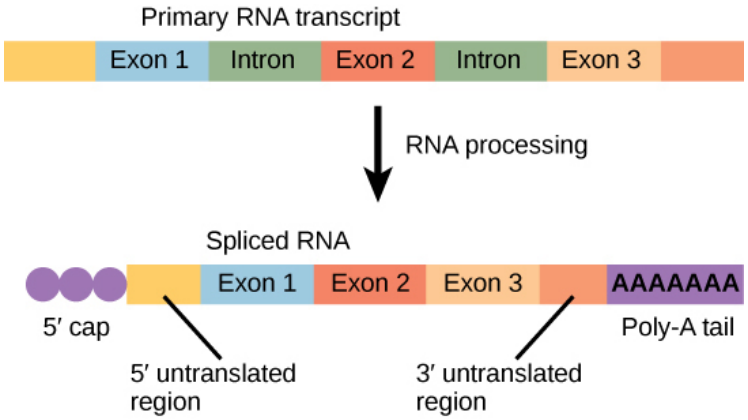


Figure 9.18 Eukaryotic mRNA contains introns that must be spliced out. A 5' cap and 3' tail are also added.

9.4 | Translation

By the end of this section, you will be able to:

- Describe the different steps in protein synthesis
- Discuss the role of ribosomes in protein synthesis
- Describe the genetic code and how the nucleotide sequence determines the amino acid and the protein sequence

The synthesis of proteins is one of a cell's most energy-consuming metabolic processes. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform a wide variety of the functions of a cell. The process of translation, or protein synthesis, involves decoding an mRNA message into a polypeptide product. Amino acids are covalently strung together in lengths ranging from approximately 50 amino acids to more than 1,000.

The Protein Synthesis Machinery

In addition to the mRNA template, many other molecules contribute to the process of translation. The composition of each component may vary across species; for instance, ribosomes may consist of different numbers of ribosomal RNAs (**rRNA**) and polypeptides depending on the organism. However, the general structures and functions of the protein synthesis machinery are comparable from bacteria to human cells. Translation requires the input of an mRNA template, ribosomes, tRNAs, and various enzymatic factors (**Figure 9.19**).

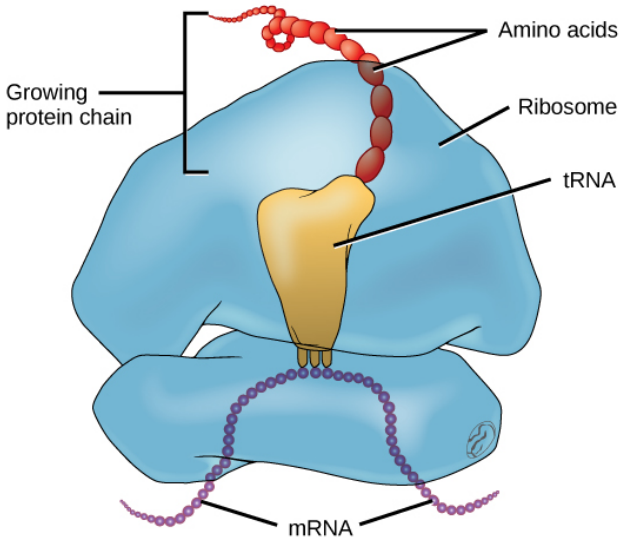


Figure 9.19 The protein synthesis machinery includes the large and small subunits of the ribosome, mRNA, and tRNA. (credit: modification of work by NIGMS, NIH)

In *E. coli*, there are 200,000 ribosomes present in every cell at any given time. A ribosome is a complex macromolecule composed of structural and catalytic rRNAs, and many distinct polypeptides. In eukaryotes, the nucleolus is completely specialized for the synthesis and assembly of rRNAs.

Ribosomes are located in the cytoplasm in prokaryotes and in the cytoplasm and endoplasmic reticulum of eukaryotes. Ribosomes are made up of a large and a small subunit that come together for translation. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds **tRNAs**, a type of RNA molecule that brings amino acids to the growing chain of the polypeptide. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction.

Depending on the species, 40 to 60 types of tRNA exist in the cytoplasm. Serving as adaptors, specific tRNAs bind to sequences on the mRNA template and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually "translate" the language of RNA into the language of proteins. For each tRNA to function, it must have its specific amino acid bonded to it. In the process of tRNA "charging," each tRNA molecule is bonded to its correct amino acid.

The Genetic Code

To summarize what we know to this point, the cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template converts nucleotide-based genetic information into a protein product. Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 letters. Each amino acid is defined by a three-nucleotide sequence called the triplet **codon**. The relationship between a nucleotide codon and its corresponding amino acid is called the **genetic code**.

Given the different numbers of "letters" in the mRNA and protein "alphabets," combinations of nucleotides corresponded to single amino acids. Using a three-nucleotide code means that there are a total of 64 (4 × 4 × 4) possible combinations; therefore, a given amino acid is encoded by more than one nucleotide triplet (**Figure 9.20**).

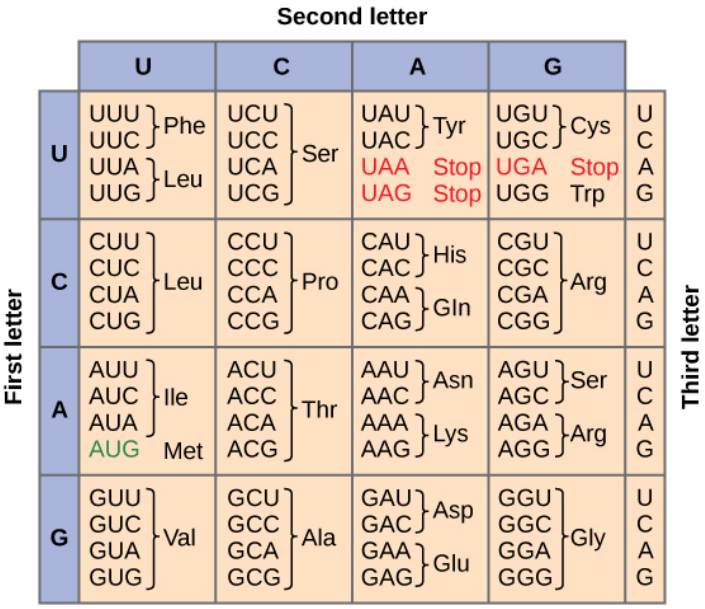


Figure 9.20 This figure shows the genetic code for translating each nucleotide triplet, or codon, in mRNA into an amino acid or a termination signal in a nascent protein. (credit: modification of work by NIH)

Three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **stop codons**. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the **start codon** to initiate translation. The reading frame for translation is set by the AUG start codon near the 5' end of the mRNA. The genetic code is universal. With a few exceptions, virtually all species use the same genetic code for protein synthesis, which is powerful evidence that all life on Earth shares a common origin.

The Mechanism of Protein Synthesis

Just as with mRNA synthesis, protein synthesis can be divided into three phases: initiation, elongation, and termination. The process of translation is similar in prokaryotes and eukaryotes. Here we will explore how translation occurs in *E. coli*, a representative prokaryote, and specify any differences between prokaryotic and eukaryotic translation.

Protein synthesis begins with the formation of an initiation complex. In *E. coli*, this complex involves the small ribosome subunit, the mRNA template, three initiation factors, and a special initiator tRNA. The initiator tRNA interacts with the AUG start codon, and links to a special form of the amino acid methionine that is typically removed from the polypeptide after translation is complete.

In prokaryotes and eukaryotes, the basics of polypeptide elongation are the same, so we will review elongation from the perspective of *E. coli*. The large ribosomal subunit of *E. coli* consists of three compartments: the A site binds incoming charged tRNAs (tRNAs with their attached specific amino acids). The P site binds charged tRNAs carrying amino acids that have formed bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The E site releases dissociated tRNAs so they can be recharged with free amino acids. The ribosome shifts one codon at a time, catalyzing each process that occurs in the three sites. With each step, a charged tRNA enters the complex, the polypeptide becomes one amino acid longer, and an uncharged tRNA departs. The energy for each bond between amino acids is derived from GTP, a molecule similar to ATP (**Figure 9.21**). Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino acid polypeptide could be translated in just 10 seconds.

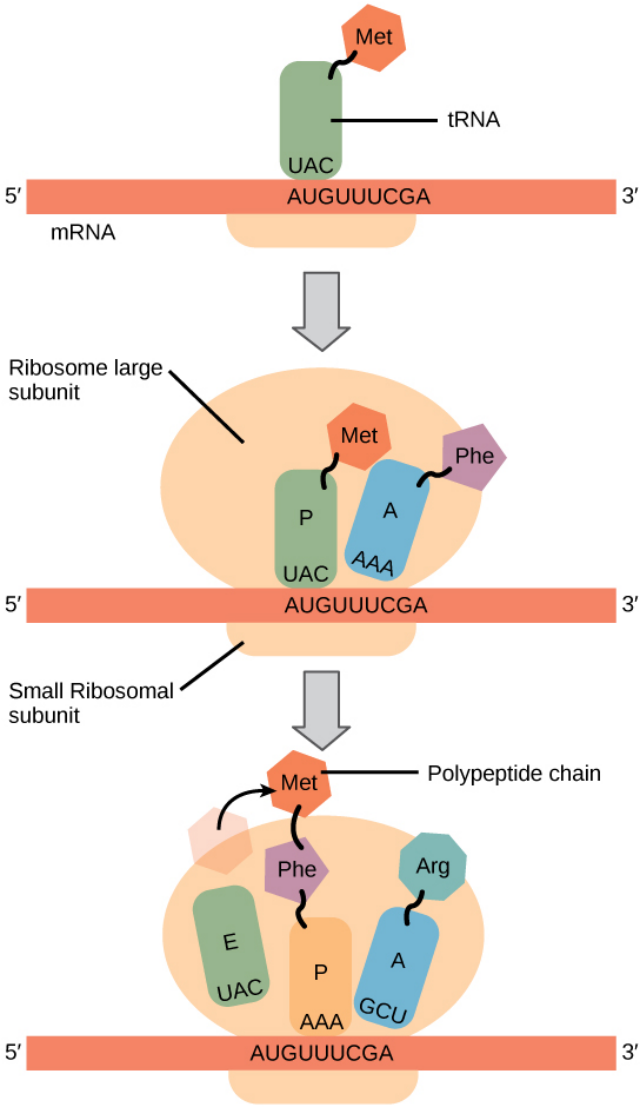


Figure 9.21 Translation begins when a tRNA anticodon recognizes a codon on the mRNA. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is encountered. When the ribosome encounters the stop codon, the growing polypeptide is released and the ribosome subunits dissociate and leave the mRNA. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

CONCEPT in ACTION

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at **this site (http://openstaxcollege.org/l/create_protein2)** .

9.5 | How Genes Are Regulated

By the end of this section, you will be able to:

- Discuss why every cell does not express all of its genes
- Describe how prokaryotic gene expression occurs at the transcriptional level
- Understand that eukaryotic gene expression occurs at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels

For a cell to function properly, necessary proteins must be synthesized at the proper time. All organisms and cells control or regulate the transcription and translation of their DNA into protein. The process of turning on a gene to produce RNA and protein is called **gene expression**. Whether in a simple unicellular organism or in a complex multicellular organism, each cell controls when and how its genes are expressed. For this to occur, there must be a mechanism to control when a gene is expressed to make RNA and protein, how much of the protein is made, and when it is time to stop making that protein because it is no longer needed.

Cells in multicellular organisms are specialized; cells in different tissues look very different and perform different functions. For example, a muscle cell is very different from a liver cell, which is very different from a skin cell. These differences are a consequence of the expression of different sets of genes in each of these cells. All cells have certain basic functions they must perform for themselves, such as converting the energy in sugar molecules into energy in ATP. Each cell also has many genes that are not expressed, and expresses many that are not expressed by other cells, such that it can carry out its specialized functions. In addition, cells will turn on or off certain genes at different times in response to changes in the environment or at different times during the development of the organism. Unicellular organisms, both eukaryotic and prokaryotic, also turn on and off genes in response to the demands of their environment so that they can respond to special conditions.

The control of gene expression is extremely complex. Malfunctions in this process are detrimental to the cell and can lead to the development of many diseases, including cancer.

Prokaryotic versus Eukaryotic Gene Expression

To understand how gene expression is regulated, we must first understand how a gene becomes a functional protein in a cell. The process occurs in both prokaryotic and eukaryotic cells, just in slightly different fashions.

Because prokaryotic organisms lack a cell nucleus, the processes of transcription and translation occur almost simultaneously. When the protein is no longer needed, transcription stops. As a result, the primary method to control what type and how much protein is expressed in a prokaryotic cell is through the regulation of DNA transcription into RNA. All the subsequent steps happen automatically. When more protein is required, more transcription occurs. Therefore, in prokaryotic cells, the control of gene expression is almost entirely at the transcriptional level.

The first example of such control was discovered using *E. coli* in the 1950s and 1960s by French researchers and is called the *lac* operon. The *lac* operon is a stretch of DNA with three adjacent genes that code for proteins that participate in the absorption and metabolism of lactose, a food source for *E. coli*. When lactose is not present in the bacterium's environment, the *lac* genes are transcribed in small amounts. When lactose is present, the genes are transcribed and the bacterium is able to use the lactose as a food source. The operon also contains a promoter sequence to which the RNA polymerase binds to begin transcription; between the promoter and the three genes is a region called the operator. When there is no lactose present, a protein known as a repressor binds to the operator and prevents RNA polymerase from binding to the promoter, except in rare cases. Thus very little of the protein products of the three genes is made. When lactose is present, an end product of lactose metabolism binds to the repressor protein and prevents it from binding to the operator. This allows RNA polymerase to bind to the promoter and freely transcribe the three genes, allowing the organism to metabolize the lactose.

Eukaryotic cells, in contrast, have intracellular organelles and are much more complex. Recall that in eukaryotic cells, the DNA is contained inside the cell's nucleus and it is transcribed into mRNA there. The newly synthesized mRNA is then transported out of the nucleus into the cytoplasm, where ribosomes translate the mRNA into protein. The processes of transcription and translation are physically separated by the nuclear membrane; transcription occurs only within the nucleus, and translation only occurs outside the nucleus in the cytoplasm. The regulation of gene expression can occur at all stages of the process (**Figure 9.22**). Regulation may occur when the DNA is uncoiled and loosened from nucleosomes to bind transcription factors (**epigenetic** level), when the RNA is transcribed (transcriptional level), when RNA is processed and

exported to the cytoplasm after it is transcribed (**post-transcriptional** level), when the RNA is translated into protein (translational level), or after the protein has been made (**post-translational** level).

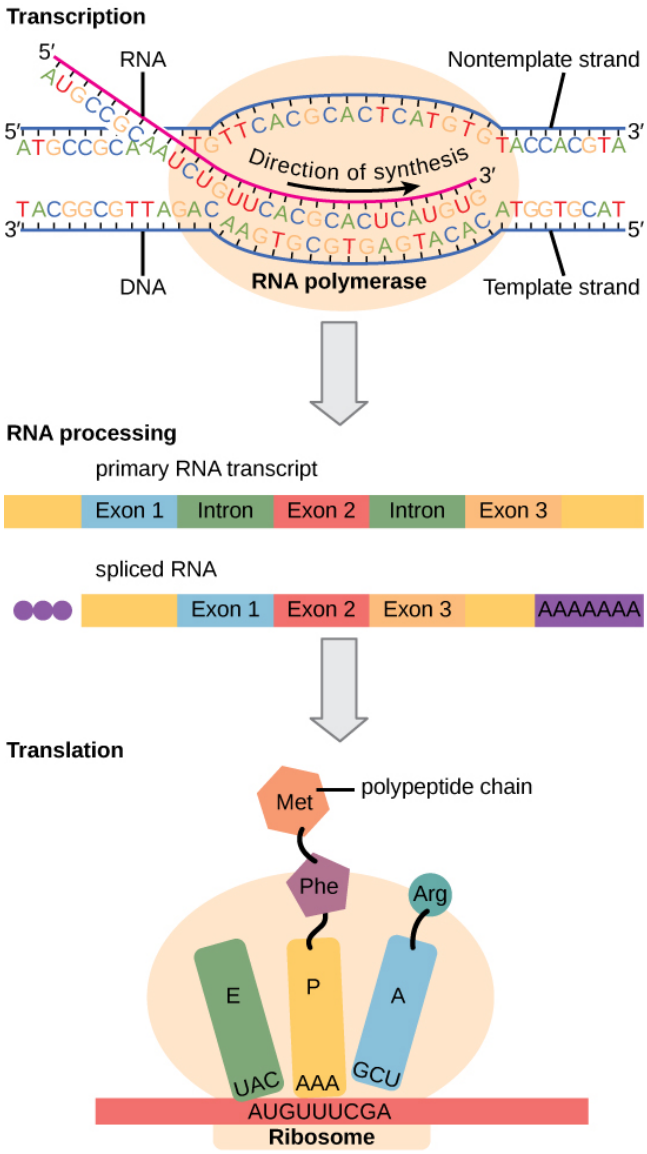


Figure 9.22 Eukaryotic gene expression is regulated during transcription and RNA processing, which take place in the nucleus, as well as during protein translation, which takes place in the cytoplasm. Further regulation may occur through post-translational modifications of proteins.

The differences in the regulation of gene expression between prokaryotes and eukaryotes are summarized in **Table 9.2**.

Differences in the Regulation of Gene Expression of Prokaryotic and Eukaryotic Organisms

Prokaryotic organisms	Eukaryotic organisms
Lack nucleus	Contain nucleus

Table 9.2

Differences in the Regulation of Gene Expression of Prokaryotic and Eukaryotic Organisms

Prokaryotic organisms	Eukaryotic organisms
RNA transcription and protein translation occur almost simultaneously	RNA transcription occurs prior to protein translation, and it takes place in the nucleus. RNA translation to protein occurs in the cytoplasm. RNA post-processing includes addition of a 5' cap, poly-A tail, and excision of introns and splicing of exons.
Gene expression is regulated primarily at the transcriptional level	Gene expression is regulated at many levels (epigenetic, transcriptional, post-transcriptional, translational, and post-translational)

Table 9.2

evolution IN ACTION

Alternative RNA Splicing

In the 1970s, genes were first observed that exhibited **alternative RNA splicing**. Alternative RNA splicing is a mechanism that allows different protein products to be produced from one gene when different combinations of introns (and sometimes exons) are removed from the transcript (Figure 9.23). This alternative splicing can be haphazard, but more often it is controlled and acts as a mechanism of gene regulation, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells, or at different stages of development. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; according to one estimate, 70% of genes in humans are expressed as multiple proteins through alternative splicing.

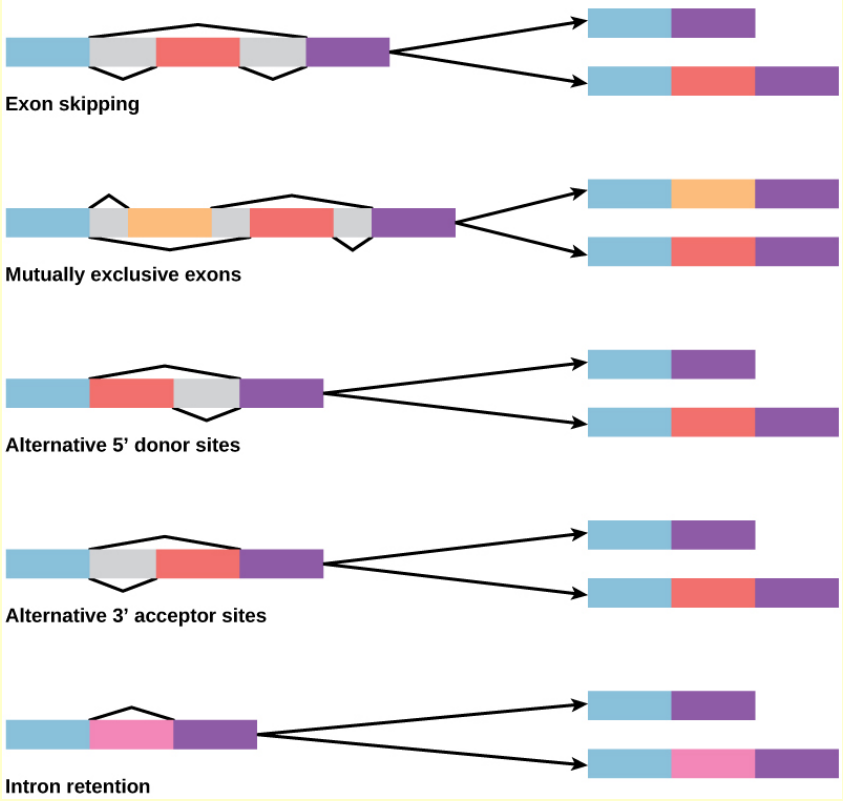


Figure 9.23 There are five basic modes of alternative splicing. Segments of pre-mRNA with exons shown in blue, red, orange, and pink can be spliced to produce a variety of new mature mRNA segments.

How could alternative splicing evolve? Introns have a beginning and ending recognition sequence, and it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and find the end of the next intron, thus removing two introns and the intervening exon. In fact, there are mechanisms in place to prevent such exon skipping, but mutations are likely to lead to their failure. Such "mistakes" would more than likely produce a nonfunctional protein. Indeed, the cause of many genetic diseases is alternative splicing rather than mutations in a sequence. However, alternative splicing would create a protein variant without the loss of the original protein, opening up possibilities for adaptation of the new variant to new functions. Gene duplication has played an important role in the evolution of new functions in a similar way—by providing genes that may evolve without eliminating the original functional protein.

KEY TERMS

alternative RNA splicing a post-transcriptional gene regulation mechanism in eukaryotes in which multiple protein products are produced by a single gene through alternative splicing combinations of the RNA transcript

codon three consecutive nucleotides in mRNA that specify the addition of a specific amino acid or the release of a polypeptide chain during translation

deoxyribose a five-carbon sugar molecule with a hydrogen atom rather than a hydroxyl group in the 2' position; the sugar component of DNA nucleotides

DNA ligase the enzyme that catalyzes the joining of DNA fragments together

DNA polymerase an enzyme that synthesizes a new strand of DNA complementary to a template strand

double helix the molecular shape of DNA in which two strands of nucleotides wind around each other in a spiral shape

epigenetic describing non-genetic regulatory factors, such as changes in modifications to histone proteins and DNA that control accessibility to genes in chromosomes

exon a sequence present in protein-coding mRNA after completion of pre-mRNA splicing

gene expression processes that control whether a gene is expressed

genetic code the amino acids that correspond to three-nucleotide codons of mRNA

helicase an enzyme that helps to open up the DNA helix during DNA replication by breaking the hydrogen bonds

intron non–protein-coding intervening sequences that are spliced from mRNA during processing

lagging strand during replication of the 3' to 5' strand, the strand that is replicated in short fragments and away from the replication fork

leading strand the strand that is synthesized continuously in the 5' to 3' direction that is synthesized in the direction of the replication fork

mismatch repair a form of DNA repair in which non-complementary nucleotides are recognized, excised, and replaced with correct nucleotides

mRNA messenger RNA; a form of RNA that carries the nucleotide sequence code for a protein sequence that is translated into a polypeptide sequence

mutation a permanent variation in the nucleotide sequence of a genome

nitrogenous base a nitrogen-containing molecule that acts as a base; often referring to one of the purine or pyrimidine components of nucleic acids

nontemplate strand the strand of DNA that is not used to transcribe mRNA; this strand is identical to the mRNA except that T nucleotides in the DNA are replaced by U nucleotides in the mRNA

nucleotide excision repair a form of DNA repair in which the DNA molecule is unwound and separated in the region of the nucleotide damage, the damaged nucleotides are removed and replaced with new nucleotides using the complementary strand, and the DNA strand is resealed and allowed to rejoin its complement

Okazaki fragments the DNA fragments that are synthesized in short stretches on the lagging strand

phosphate group a molecular group consisting of a central phosphorus atom bound to four oxygen atoms

post-transcriptional control of gene expression after the RNA molecule has been created but before it is translated into protein

post-translational control of gene expression after a protein has been created

primer a short stretch of RNA nucleotides that is required to initiate replication and allow DNA polymerase to bind and begin replication

promoter a sequence on DNA to which RNA polymerase and associated factors bind and initiate transcription

replication fork the Y-shaped structure formed during the initiation of replication

RNA polymerase an enzyme that synthesizes an RNA strand from a DNA template strand

rRNA ribosomal RNA; molecules of RNA that combine to form part of the ribosome

semiconservative replication the method used to replicate DNA in which the double-stranded molecule is separated and each strand acts as a template for a new strand to be synthesized, so the resulting DNA molecules are composed of one new strand of nucleotides and one old strand of nucleotides

splicing the process of removing introns and reconnecting exons in a pre-mRNA

start codon the AUG (or, rarely GUG) on an mRNA from which translation begins; always specifies methionine

stop codon one of the three mRNA codons that specifies termination of translation

telomerase an enzyme that contains a catalytic part and an inbuilt RNA template; it functions to maintain telomeres at chromosome ends

telomere the DNA at the end of linear chromosomes

template strand the strand of DNA that specifies the complementary mRNA molecule

transcription bubble the region of locally unwound DNA that allows for transcription of mRNA

tRNA transfer RNA; an RNA molecule that contains a specific three-nucleotide anticodon sequence to pair with the mRNA codon and also binds to a specific amino acid

CHAPTER SUMMARY

9.1 The Structure of DNA

The model of the double-helix structure of DNA was proposed by Watson and Crick. The DNA molecule is a polymer of nucleotides. Each nucleotide is composed of a nitrogenous base, a five-carbon sugar (deoxyribose), and a phosphate group. There are four nitrogenous bases in DNA, two purines (adenine and guanine) and two pyrimidines (cytosine and thymine). A DNA molecule is composed of two strands. Each strand is composed of nucleotides bonded together covalently between the phosphate group of one and the deoxyribose sugar of the next. From this backbone extend the bases. The bases of one strand bond to the bases of the second strand with hydrogen bonds. Adenine always bonds with thymine, and cytosine always bonds with guanine. The bonding causes the two strands to spiral around each other in a shape called a double helix. Ribonucleic acid (RNA) is a second nucleic acid found in cells. RNA is a single-stranded polymer of nucleotides. It also differs from DNA in that it contains the sugar ribose, rather than deoxyribose, and the nucleotide uracil rather than thymine. Various RNA molecules function in the process of forming proteins from the genetic code in DNA.

Prokaryotes contain a single, double-stranded circular chromosome. Eukaryotes contain double-stranded linear DNA molecules packaged into chromosomes. The DNA helix is wrapped around proteins to form nucleosomes. The protein coils are further coiled, and during mitosis and meiosis, the chromosomes become even more greatly coiled to facilitate their movement. Chromosomes have two distinct regions which can be distinguished by staining, reflecting different degrees of packaging and determined by whether the DNA in a region is being expressed (euchromatin) or not (heterochromatin).

9.2 DNA Replication

DNA replicates by a semi-conservative method in which each of the two parental DNA strands act as a template for new DNA to be synthesized. After replication, each DNA has one parental or “old” strand, and one daughter or “new” strand.

Replication in eukaryotes starts at multiple origins of replication, while replication in prokaryotes starts from a single origin of replication. The DNA is opened with enzymes, resulting in the formation of the replication fork. Primase synthesizes an RNA primer to initiate synthesis by DNA polymerase, which can add nucleotides in only one direction.

One strand is synthesized continuously in the direction of the replication fork; this is called the leading strand. The other strand is synthesized in a direction away from the replication fork, in short stretches of DNA known as Okazaki fragments. This strand is known as the lagging strand. Once replication is completed, the RNA primers are replaced by DNA nucleotides and the DNA is sealed with DNA ligase.

The ends of eukaryotic chromosomes pose a problem, as polymerase is unable to extend them without a primer. Telomerase, an enzyme with an inbuilt RNA template, extends the ends by copying the RNA template and extending one end of the chromosome. DNA polymerase can then extend the DNA using the primer. In this way, the ends of the chromosomes are protected. Cells have mechanisms for repairing DNA when it becomes damaged or errors are made in replication. These mechanisms include mismatch repair to replace nucleotides that are paired with a non-complementary base and nucleotide excision repair, which removes bases that are damaged such as thymine dimers.

9.3 Transcription

In prokaryotes, mRNA synthesis is initiated at a promoter sequence on the DNA template. Elongation synthesizes new mRNA. Termination liberates the mRNA and occurs by mechanisms that stall the RNA polymerase and cause it to fall off the DNA template. Newly transcribed eukaryotic mRNAs are modified with a cap and a poly-A tail. These structures protect the mature mRNA from degradation and help export it from the nucleus. Eukaryotic mRNAs also undergo splicing, in which introns are removed and exons are reconnected with single-nucleotide accuracy. Only finished mRNAs are exported from the nucleus to the cytoplasm.

9.4 Translation

The central dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is the correspondence between the three-nucleotide mRNA codon and an amino acid. The genetic code is “translated” by the tRNA molecules, which associate a specific codon with a specific amino acid. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three stop codons. This means that more than one codon corresponds to an amino acid. Almost every species on the planet uses the same genetic code.

The players in translation include the mRNA template, ribosomes, tRNAs, and various enzymatic factors. The small ribosomal subunit binds to the mRNA template. Translation begins at the initiating AUG on the mRNA. The formation of bonds occurs between sequential amino acids specified by the mRNA template according to the genetic code. The ribosome accepts charged tRNAs, and as it steps along the mRNA, it catalyzes bonding between the new amino acid and the end of the growing polypeptide. The entire mRNA is translated in three-nucleotide “steps” of the ribosome. When a stop codon is encountered, a release factor binds and dissociates the components and frees the new protein.

9.5 How Genes Are Regulated

While all somatic cells within an organism contain the same DNA, not all cells within that organism express the same proteins. Prokaryotic organisms express the entire DNA they encode in every cell, but not necessarily all at the same time. Proteins are expressed only when they are needed. Eukaryotic organisms express a subset of the DNA that is encoded in any given cell. In each cell type, the type and amount of protein is regulated by controlling gene expression. To express a protein, the DNA is first transcribed into RNA, which is then translated into proteins. In prokaryotic cells, these processes occur almost simultaneously. In eukaryotic cells, transcription occurs in the nucleus and is separate from the translation that occurs in the cytoplasm. Gene expression in prokaryotes is regulated only at the transcriptional level, whereas in eukaryotic cells, gene expression is regulated at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels.

ART CONNECTION QUESTIONS

1. **Figure 9.10** You isolate a cell strain in which the joining together of Okazaki fragments is impaired and suspect that a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

REVIEW QUESTIONS

2. Which of the following does cytosine pair with?

a. guanine
b. thymine
c. adenine
d. a pyrimidine

3. Prokaryotes contain a _____chromosome, and eukaryotes contain _____ chromosomes.
- a. single-stranded circular; single-stranded linear
- b. single-stranded linear; single-stranded circular
- c. double-stranded circular; double-stranded linear
- d. double-stranded linear; double-stranded circular

4. DNA replicates by which of the following models?

- a. conservative
- b. semiconservative
- c. dispersive
- d. none of the above

5. The initial mechanism for repairing nucleotide errors in DNA is _____.
- a. mismatch repair
- b. DNA polymerase proofreading
- c. nucleotide excision repair
- d. thymine dimers

6. A promoter is _____.
- a. a specific sequence of DNA nucleotides
- b. a specific sequence of RNA nucleotides
- c. a protein that binds to DNA
- d. an enzyme that synthesizes RNA

7. Portions of eukaryotic mRNA sequence that are removed during RNA processing are _____.
- a. exons
- b. caps
- c. poly-A tails
- d. introns

8. The RNA components of ribosomes are synthesized in the _____.
- a. cytoplasm
- b. nucleus
- c. nucleolus
- d. endoplasmic reticulum

9. How long would the peptide be that is translated from this MRNA sequence: 5'-AUGGGCUACCGA-3'?

- a. 0
- b. 2
- c. 3
- d. 4

10. Control of gene expression in eukaryotic cells occurs at which level(s)?
- a. only the transcriptional level
- b. epigenetic and transcriptional levels
- c. epigenetic, transcriptional, and translational levels
- d. epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels

11. Post-translational control refers to:
- a. regulation of gene expression after transcription
- b. regulation of gene expression after translation
- c. control of epigenetic activation
- d. period between transcription and translation

CRITICAL THINKING QUESTIONS

12. Describe the organization of the eukaryotic chromosome.

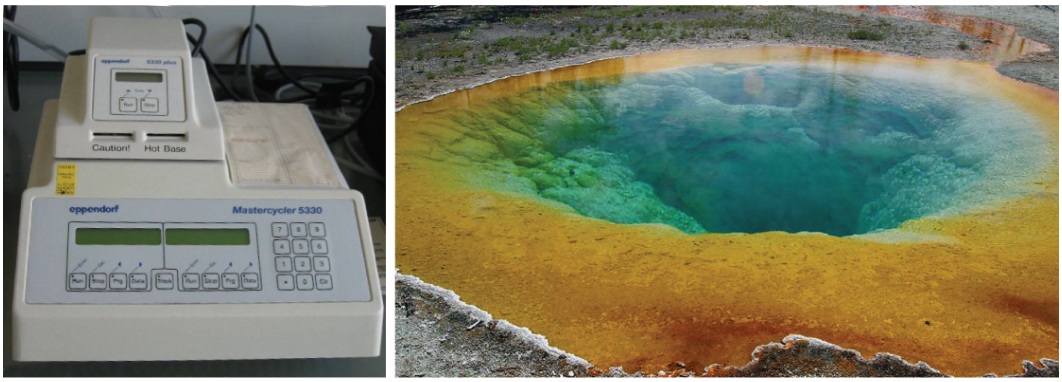
13. Describe the structure and complementary base pairing of DNA.

14. How do the linear chromosomes in eukaryotes ensure that its ends are replicated completely?

15. Transcribe and translate the following DNA sequence (nontemplate strand): 5'-ATGGCCGGTTATTAAGCA-3'

16. Describe how controlling gene expression will alter the overall protein levels in the cell.

10 | BIOTECHNOLOGY



(a) (b)

Figure 10.1 (a) A thermal cycler, such as the one shown here, is a basic tool used to study DNA in a process called the polymerase chain reaction (PCR). The polymerase enzyme most often used with PCR comes from a strain of bacteria that lives in (b) the hot springs of Yellowstone National Park. (credit a: modification of work by Magnus Manske; credit b: modification of work by Jon Sullivan)

Chapter Outline

Introduction

The latter half of the twentieth century began with the discovery of the structure of DNA, then progressed to the development of the basic tools used to study and manipulate DNA. These advances, as well as advances in our understanding of and ability to manipulate cells, have led some to refer to the twenty-first century as the biotechnology century. The rate of discovery and of the development of new applications in medicine, agriculture, and energy is expected to accelerate, bringing huge benefits to humankind and perhaps also significant risks. Many of these developments are expected to raise significant ethical and social questions that human societies have not yet had to consider.

10.1 | Cloning and Genetic Engineering

By the end of this section, you will be able to:

- Explain the basic techniques used to manipulate genetic material
- Explain molecular and reproductive cloning

Biotechnology is the use of artificial methods to modify the genetic material of living organisms or cells to produce novel compounds or to perform new functions. Biotechnology has been used for improving livestock and crops since the beginning of agriculture through selective breeding. Since the discovery of the structure of DNA in 1953, and particularly since the development of tools and methods to manipulate DNA in the 1970s, biotechnology has become synonymous with the manipulation of organisms' DNA at the molecular level. The primary applications of this technology are in medicine (for the production of vaccines and antibiotics) and in agriculture (for the genetic modification of crops). Biotechnology also has many industrial applications, such as fermentation, the treatment of oil spills, and the production of biofuels, as well as many household applications such as the use of enzymes in laundry detergent.

Manipulating Genetic Material

To accomplish the applications described above, biotechnologists must be able to extract, manipulate, and analyze nucleic acids.

Review of Nucleic Acid Structure

To understand the basic techniques used to work with nucleic acids, remember that nucleic acids are macromolecules made of nucleotides (a sugar, a phosphate, and a nitrogenous base). The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus of eukaryotic organisms is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases.

Unlike DNA in eukaryotic cells, RNA molecules leave the nucleus. Messenger RNA (mRNA) is analyzed most frequently because it represents the protein-coding genes that are being expressed in the cell.

Isolation of Nucleic Acids

To study or manipulate nucleic acids, the DNA must first be extracted from cells. Various techniques are used to extract different types of DNA (**Figure 10.2**). Most nucleic acid extraction techniques involve steps to break open the cell, and then the use of enzymatic reactions to destroy all undesired macromolecules. Cells are broken open using a detergent solution containing buffering compounds. To prevent degradation and contamination, macromolecules such as proteins and RNA are inactivated using enzymes. The DNA is then brought out of solution using alcohol. The resulting DNA, because it is made up of long polymers, forms a gelatinous mass.

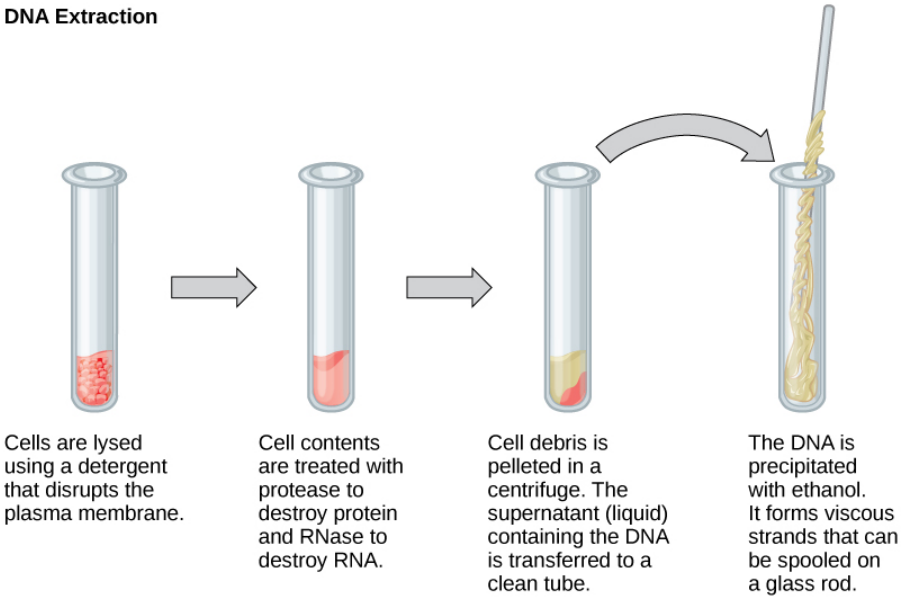


Figure 10.2 This diagram shows the basic method used for the extraction of DNA.

RNA is studied to understand gene expression patterns in cells. RNA is naturally very unstable because enzymes that break down RNA are commonly present in nature. Some are even secreted by our own skin and are very difficult to inactivate. Similar to DNA extraction, RNA extraction involves the use of various buffers and enzymes to inactivate other macromolecules and preserve only the RNA.

Gel Electrophoresis

Because nucleic acids are negatively charged ions at neutral or alkaline pH in an aqueous environment, they can be moved by an electric field. **Gel electrophoresis** is a technique used to separate charged molecules on the basis of size and charge. The nucleic acids can be separated as whole chromosomes or as fragments. The nucleic acids are loaded into a slot at one end of a gel matrix, an electric current is applied, and negatively charged molecules are pulled toward the opposite end of the gel (the end with the positive electrode). Smaller molecules move through the pores in the gel faster than larger molecules; this difference in the rate of migration separates the fragments on the basis of size. The nucleic acids in a gel matrix are invisible until they are stained with a compound that allows them to be seen, such as a dye. Distinct fragments of nucleic acids appear as bands at specific distances from the top of the gel (the negative electrode end) that are based on their size (**Figure 10.3**). A mixture of many fragments of varying sizes appear as a long smear, whereas uncut genomic DNA is usually too large to run through the gel and forms a single large band at the top of the gel.

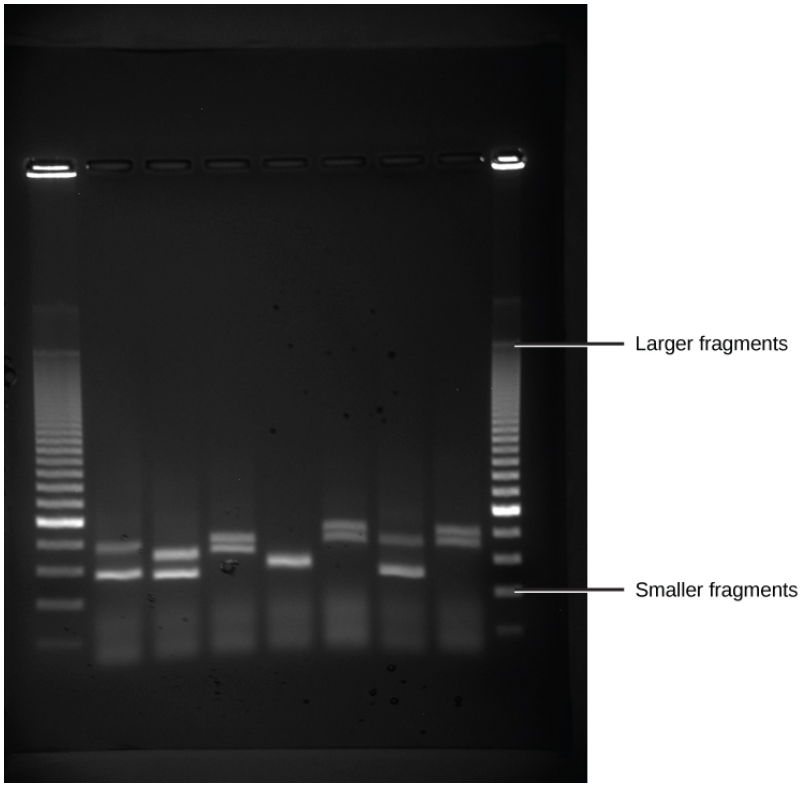


Figure 10.3 Shown are DNA fragments from six samples run on a gel, stained with a fluorescent dye and viewed under UV light. (credit: modification of work by James Jacob, Tompkins Cortland Community College)

Polymerase Chain Reaction

DNA analysis often requires focusing on one or more specific regions of the genome. It also frequently involves situations in which only one or a few copies of a DNA molecule are available for further analysis. These amounts are insufficient for most procedures, such as gel electrophoresis. **Polymerase chain reaction (PCR)** is a technique used to rapidly increase the number of copies of specific regions of DNA for further analyses (**Figure 10.4**). PCR uses a special form of DNA polymerase, the enzyme that replicates DNA, and other short nucleotide sequences called primers that base pair to a specific portion of the DNA being replicated. PCR is used for many purposes in laboratories. These include: 1) the identification of the owner of a DNA sample left at a crime scene; 2) paternity analysis; 3) the comparison of small amounts of ancient DNA with modern organisms; and 4) determining the sequence of nucleotides in a specific region.

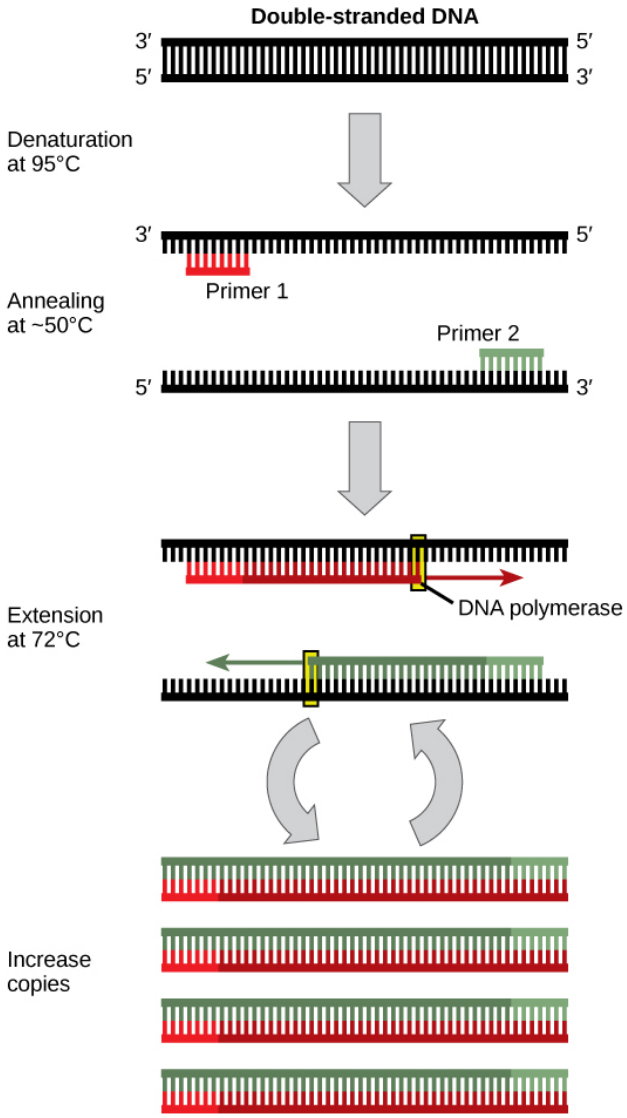


Figure 10.4 Polymerase chain reaction, or PCR, is used to produce many copies of a specific sequence of DNA using a special form of DNA polymerase.

Cloning

In general, **cloning** means the creation of a perfect replica. Typically, the word is used to describe the creation of a genetically identical copy. In biology, the re-creation of a whole organism is referred to as “reproductive cloning.” Long before attempts were made to clone an entire organism, researchers learned how to copy short stretches of DNA—a process that is referred to as molecular cloning.

Molecular Cloning

Cloning allows for the creation of multiple copies of genes, expression of genes, and study of specific genes. To get the DNA fragment into a bacterial cell in a form that will be copied or expressed, the fragment is first inserted into a plasmid. A **plasmid** (also called a vector in this context) is a small circular DNA molecule that replicates independently of the chromosomal DNA in bacteria. In cloning, the plasmid molecules can be used to provide a "vehicle" in which to insert a desired DNA fragment. Modified plasmids are usually reintroduced into a bacterial host for replication. As the bacteria divide, they copy their own DNA (including the plasmids). The inserted DNA fragment is copied along with the rest of the bacterial DNA. In a bacterial cell, the fragment of DNA from the human genome (or another organism that is being studied) is referred to as foreign DNA to differentiate it from the DNA of the bacterium (the host DNA).

Plasmids occur naturally in bacterial populations (such as *Escherichia coli*) and have genes that can contribute favorable traits to the organism, such as antibiotic resistance (the ability to be unaffected by antibiotics). Plasmids have been highly engineered as vectors for molecular cloning and for the subsequent large-scale production of important molecules, such as

insulin. A valuable characteristic of plasmid vectors is the ease with which a foreign DNA fragment can be introduced. These plasmid vectors contain many short DNA sequences that can be cut with different commonly available **restriction enzymes**. Restriction enzymes (also called restriction endonucleases) recognize specific DNA sequences and cut them in a predictable manner; they are naturally produced by bacteria as a defense mechanism against foreign DNA. Many restriction enzymes make staggered cuts in the two strands of DNA, such that the cut ends have a 2- to 4-nucleotide single-stranded overhang. The sequence that is recognized by the restriction enzyme is a four- to eight-nucleotide sequence that is a palindrome. Like with a word palindrome, this means the sequence reads the same forward and backward. In most cases, the sequence reads the same forward on one strand and backward on the complementary strand. When a staggered cut is made in a sequence like this, the overhangs are complementary (**Figure 10.5**).

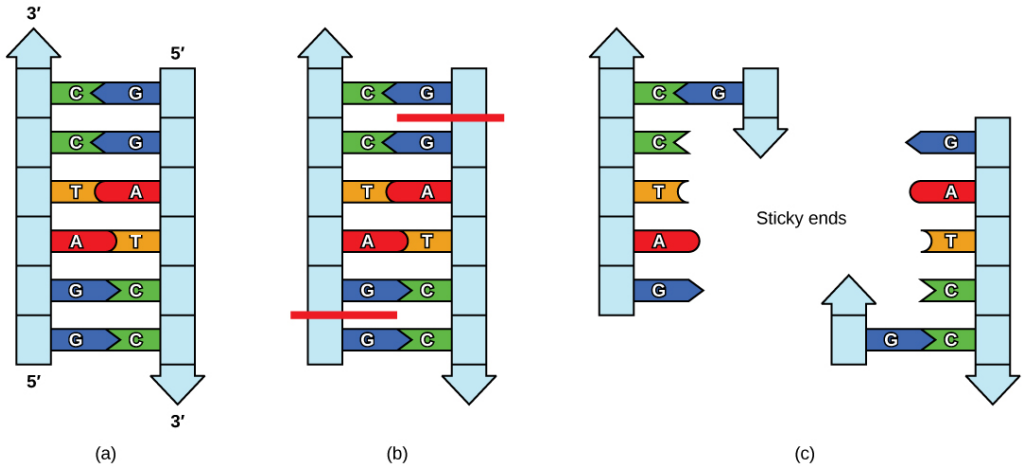


Figure 10.5 In this (a) six-nucleotide restriction enzyme recognition site, notice that the sequence of six nucleotides reads the same in the 5' to 3' direction on one strand as it does in the 5' to 3' direction on the complementary strand. This is known as a palindrome. (b) The restriction enzyme makes breaks in the DNA strands, and (c) the cut in the DNA results in “sticky ends”. Another piece of DNA cut on either end by the same restriction enzyme could attach to these sticky ends and be inserted into the gap made by this cut.

Because these overhangs are capable of coming back together by hydrogen bonding with complementary overhangs on a piece of DNA cut with the same restriction enzyme, these are called “sticky ends.” The process of forming hydrogen bonds between complementary sequences on single strands to form double-stranded DNA is called **annealing**. Addition of an enzyme called DNA ligase, which takes part in DNA replication in cells, permanently joins the DNA fragments when the sticky ends come together. In this way, any DNA fragment can be spliced between the two ends of a plasmid DNA that has been cut with the same restriction enzyme (**Figure 10.6**).

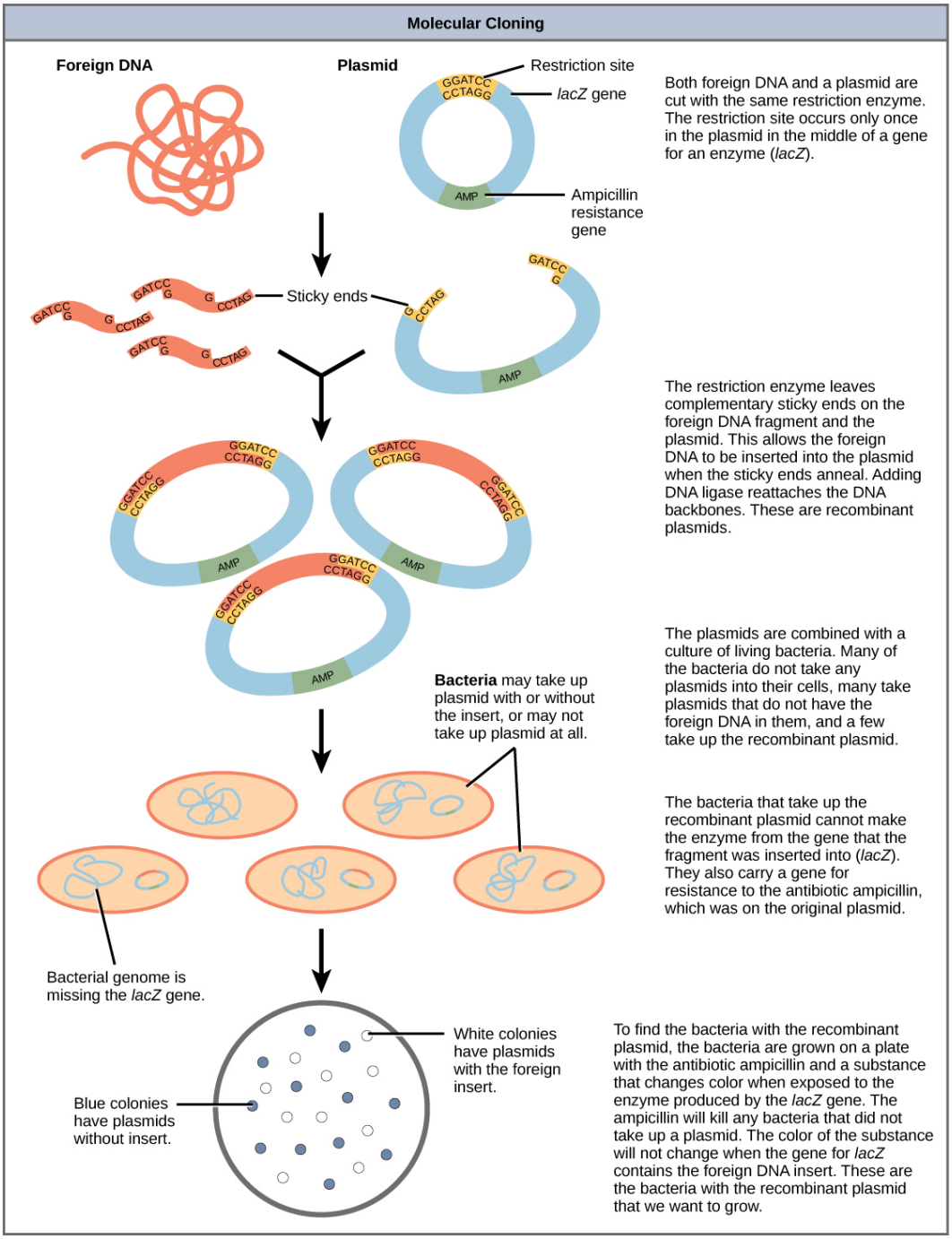


Figure 10.6 This diagram shows the steps involved in molecular cloning.

Plasmids with foreign DNA inserted into them are called **recombinant DNA** molecules because they contain new combinations of genetic material. Proteins that are produced from recombinant DNA molecules are called **recombinant proteins**. Not all recombinant plasmids are capable of expressing genes. Plasmids may also be engineered to express proteins only when stimulated by certain environmental factors, so that scientists can control the expression of the recombinant proteins.

Reproductive Cloning

Reproductive cloning is a method used to make a clone or an identical copy of an entire multicellular organism. Most multicellular organisms undergo reproduction by sexual means, which involves the contribution of DNA from two individuals (parents), making it impossible to generate an identical copy or a clone of either parent. Recent advances in biotechnology have made it possible to reproductively clone mammals in the laboratory.

Natural sexual reproduction involves the union, during fertilization, of a sperm and an egg. Each of these gametes is haploid, meaning they contain one set of chromosomes in their nuclei. The resulting cell, or zygote, is then diploid and contains two sets of chromosomes. This cell divides mitotically to produce a multicellular organism. However, the union of just any two cells cannot produce a viable zygote; there are components in the cytoplasm of the egg cell that are essential for the early development of the embryo during its first few cell divisions. Without these provisions, there would be no subsequent development. Therefore, to produce a new individual, both a diploid genetic complement and an egg cytoplasm are required. The approach to producing an artificially cloned individual is to take the egg cell of one individual and to remove the haploid nucleus. Then a diploid nucleus from a body cell of a second individual, the donor, is put into the egg cell. The egg is then stimulated to divide so that development proceeds. This sounds simple, but in fact it takes many attempts before each of the steps is completed successfully.

The first cloned agricultural animal was Dolly, a sheep who was born in 1996. The success rate of reproductive cloning at the time was very low. Dolly lived for six years and died of a lung tumor (**Figure 10.7**). There was speculation that because the cell DNA that gave rise to Dolly came from an older individual, the age of the DNA may have affected her life expectancy. Since Dolly, several species of animals (such as horses, bulls, and goats) have been successfully cloned.

There have been attempts at producing cloned human embryos as sources of embryonic stem cells. In the procedure, the DNA from an adult human is introduced into a human egg cell, which is then stimulated to divide. The technology is similar to the technology that was used to produce Dolly, but the embryo is never implanted into a surrogate mother. The cells produced are called embryonic stem cells because they have the capacity to develop into many different kinds of cells, such as muscle or nerve cells. The stem cells could be used to research and ultimately provide therapeutic applications, such as replacing damaged tissues. The benefit of cloning in this instance is that the cells used to regenerate new tissues would be a perfect match to the donor of the original DNA. For example, a leukemia patient would not require a sibling with a tissue match for a bone-marrow transplant.

art CONNECTION

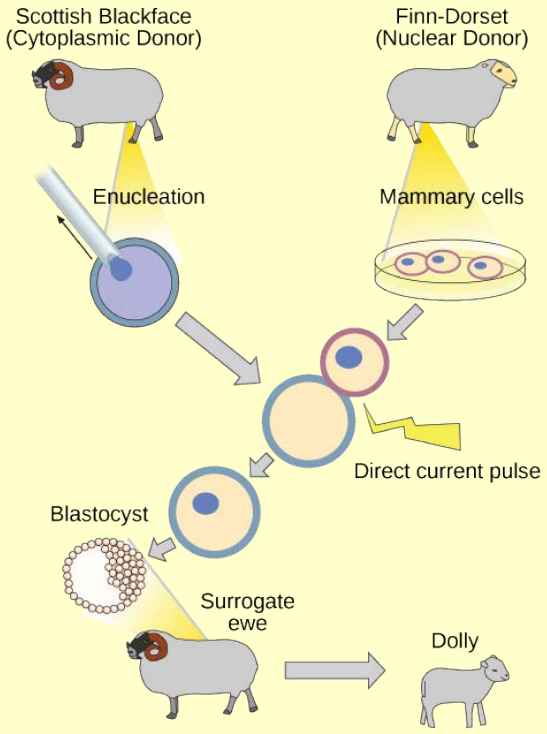


Figure 10.7 Dolly the sheep was the first agricultural animal to be cloned. To create Dolly, the nucleus was removed from a donor egg cell. The enucleated egg was placed next to the other cell, then they were shocked to fuse. They were shocked again to start division. The cells were allowed to divide for several days until an early embryonic stage was reached, before being implanted in a surrogate mother.

Why was Dolly a Finn-Dorset and not a Scottish Blackface sheep?

Genetic Engineering

Using recombinant DNA technology to modify an organism's DNA to achieve desirable traits is called **genetic engineering**. Addition of foreign DNA in the form of recombinant DNA vectors that are generated by molecular cloning is the most common method of genetic engineering. An organism that receives the recombinant DNA is called a **genetically modified organism** (GMO). If the foreign DNA that is introduced comes from a different species, the host organism is called **transgenic**. Bacteria, plants, and animals have been genetically modified since the early 1970s for academic, medical, agricultural, and industrial purposes. These applications will be examined in more detail in the next module.

Watch this **short video (http://openstaxcollege.org/l/transgenic)** explaining how scientists create a transgenic animal.

Although the classic methods of studying the function of genes began with a given phenotype and determined the genetic basis of that phenotype, modern techniques allow researchers to start at the DNA sequence level and ask: "What does this gene or DNA element do?" This technique, called **reverse genetics**, has resulted in reversing the classical genetic methodology. One example of this method is analogous to damaging a body part to determine its function. An insect that loses a wing cannot fly, which means that the wing's function is flight. The classic genetic method compares insects that cannot fly with insects that can fly, and observes that the non-flying insects have lost wings. Similarly in a reverse genetics approach, mutating or deleting genes provides researchers with clues about gene function. Alternately, reverse genetics can be used to cause a gene to overexpress itself to determine what phenotypic effects may occur.

10.2 | Biotechnology in Medicine and Agriculture

By the end of this section, you will be able to:

- Describe uses of biotechnology in medicine
- Describe uses of biotechnology in agriculture

It is easy to see how biotechnology can be used for medicinal purposes. Knowledge of the genetic makeup of our species, the genetic basis of heritable diseases, and the invention of technology to manipulate and fix mutant genes provides methods to treat diseases. Biotechnology in agriculture can enhance resistance to disease, pests, and environmental stress to improve both crop yield and quality.

Genetic Diagnosis and Gene Therapy

The process of testing for suspected genetic defects before administering treatment is called genetic diagnosis by genetic testing. In some cases in which a genetic disease is present in an individual's family, family members may be advised to undergo genetic testing. For example, mutations in the *BRCA* genes may increase the likelihood of developing breast and ovarian cancers in women and some other cancers in women and men. A woman with breast cancer can be screened for these mutations. If one of the high-risk mutations is found, her female relatives may also wish to be screened for that particular mutation, or simply be more vigilant for the occurrence of cancers. Genetic testing is also offered for fetuses (or embryos with in vitro fertilization) to determine the presence or absence of disease-causing genes in families with specific debilitating diseases.

CONCEPT in ACTION

See how **human DNA is extracted (http://openstaxcollege.org/l/DNA_extraction)** for uses such as genetic testing.

Gene therapy is a genetic engineering technique that may one day be used to cure certain genetic diseases. In its simplest form, it involves the introduction of a non-mutated gene at a random location in the genome to cure a disease by replacing a protein that may be absent in these individuals because of a genetic mutation. The non-mutated gene is usually introduced into diseased cells as part of a vector transmitted by a virus, such as an adenovirus, that can infect the host cell and deliver the foreign DNA into the genome of the targeted cell (**Figure 10.8**). To date, gene therapies have been primarily experimental procedures in humans. A few of these experimental treatments have been successful, but the methods may be important in the future as the factors limiting its success are resolved.

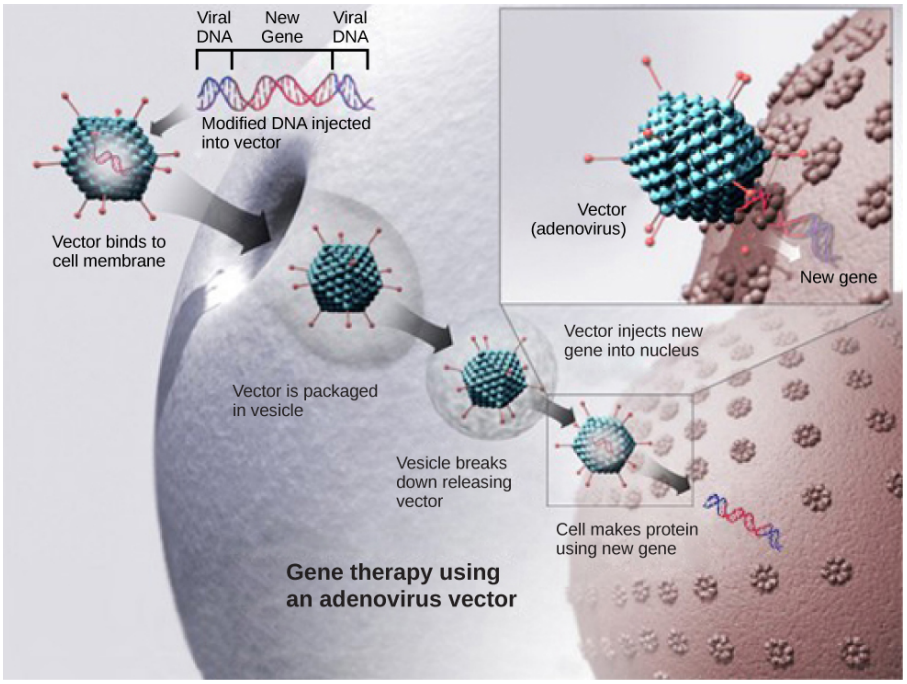


Figure 10.8 This diagram shows the steps involved in curing disease with gene therapy using an adenovirus vector. (credit: modification of work by NIH)

Production of Vaccines, Antibiotics, and Hormones

Traditional vaccination strategies use weakened or inactive forms of microorganisms or viruses to stimulate the immune system. Modern techniques use specific genes of microorganisms cloned into vectors and mass-produced in bacteria to make large quantities of specific substances to stimulate the immune system. The substance is then used as a vaccine. In some cases, such as the H1N1 flu vaccine, genes cloned from the virus have been used to combat the constantly changing strains of this virus.

Antibiotics kill bacteria and are naturally produced by microorganisms such as fungi; penicillin is perhaps the most well-known example. Antibiotics are produced on a large scale by cultivating and manipulating fungal cells. The fungal cells have typically been genetically modified to improve the yields of the antibiotic compound.

Recombinant DNA technology was used to produce large-scale quantities of the human hormone insulin in *E. coli* as early as 1978. Previously, it was only possible to treat diabetes with pig insulin, which caused allergic reactions in many humans because of differences in the insulin molecule. In addition, human growth hormone (HGH) is used to treat growth disorders

in children. The HGH gene was cloned from a cDNA (complementary DNA) library and inserted into *E. coli* cells by cloning it into a bacterial vector.

Transgenic Animals

Although several recombinant proteins used in medicine are successfully produced in bacteria, some proteins need a eukaryotic animal host for proper processing. For this reason, genes have been cloned and expressed in animals such as sheep, goats, chickens, and mice. Animals that have been modified to express recombinant DNA are called transgenic animals (**Figure 10.9**).

Figure 10.9 It can be seen that two of these mice are transgenic because they have a gene that causes them to fluoresce under a UV light. The non-transgenic mouse does not have the gene that causes fluorescence. (credit: Ingrid Moen et al.)

Several human proteins are expressed in the milk of transgenic sheep and goats. In one commercial example, the FDA has approved a blood anticoagulant protein that is produced in the milk of transgenic goats for use in humans. Mice have been used extensively for expressing and studying the effects of recombinant genes and mutations.

Transgenic Plants

Manipulating the DNA of plants (creating genetically modified organisms, or GMOs) has helped to create desirable traits such as disease resistance, herbicide, and pest resistance, better nutritional value, and better shelf life (**Figure 10.10**). Plants are the most important source of food for the human population. Farmers developed ways to select for plant varieties with desirable traits long before modern-day biotechnology practices were established.

Figure 10.10 Corn, a major agricultural crop used to create products for a variety of industries, is often modified through plant biotechnology. (credit: Keith Weller, USDA)

Transgenic plants have received DNA from other species. Because they contain unique combinations of genes and are not restricted to the laboratory, transgenic plants and other GMOs are closely monitored by government agencies to ensure that they are fit for human consumption and do not endanger other plant and animal life. Because foreign genes can spread to other species in the environment, particularly in the pollen and seeds of plants, extensive testing is required to ensure ecological stability. Staples like corn, potatoes, and tomatoes were the first crop plants to be genetically engineered.

Transformation of Plants Using Agrobacterium tumefaciens

In plants, tumors caused by the bacterium *Agrobacterium tumefaciens* occur by transfer of DNA from the bacterium to the plant. The artificial introduction of DNA into plant cells is more challenging than in animal cells because of the thick plant cell wall. Researchers used the natural transfer of DNA from *Agrobacterium* to a plant host to introduce DNA fragments of their choice into plant hosts. In nature, the disease-causing *A. tumefaciens* have a set of plasmids that contain genes that integrate into the infected plant cell's genome. Researchers manipulate the plasmids to carry the desired DNA fragment and insert it into the plant genome.

The Organic Insecticide Bacillus thuringiensis

Bacillus thuringiensis (Bt) is a bacterium that produces protein crystals that are toxic to many insect species that feed on plants. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin is activated in the intestines of the insects, death occurs within a couple of days. The crystal toxin genes have been cloned from the bacterium and introduced into plants, therefore allowing plants to produce their own crystal Bt toxin that acts against insects. Bt toxin is safe for the environment and non-toxic to mammals (including humans). As a result, it has been approved for use by organic farmers as a natural insecticide. There is some concern, however, that insects may evolve resistance to the Bt toxin in the same way that bacteria evolve resistance to antibiotics.

FlavrSavr Tomato

The first GM crop to be introduced into the market was the FlavrSavr Tomato produced in 1994. Molecular genetic technology was used to slow down the process of softening and rotting caused by fungal infections, which led to increased shelf life of the GM tomatoes. Additional genetic modification improved the flavor of this tomato. The FlavrSavr tomato did not successfully stay in the market because of problems maintaining and shipping the crop.

10.3 | Genomics and Proteomics

By the end of this section, you will be able to:

- Define genomics and proteomics
- Define whole genome sequencing
- Explain different applications of genomics and proteomics

The study of nucleic acids began with the discovery of DNA, progressed to the study of genes and small fragments, and has now exploded to the field of **genomics**. Genomics is the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species. The advances in genomics have been made possible by DNA sequencing technology. Just as information technology has led to Google Maps that enable us to get detailed information about locations around the globe, genomic information is used to create similar maps of the DNA of different organisms.

Mapping Genomes

Genome mapping is the process of finding the location of genes on each chromosome. The maps that are created are comparable to the maps that we use to navigate streets. A **genetic map** is an illustration that lists genes and their location on a chromosome. Genetic maps provide the big picture (similar to a map of interstate highways) and use genetic markers (similar to landmarks). A genetic marker is a gene or sequence on a chromosome that shows genetic linkage with a trait of interest. The genetic marker tends to be inherited with the gene of interest, and one measure of distance between them is the recombination frequency during meiosis. Early geneticists called this linkage analysis.

Physical maps get into the intimate details of smaller regions of the chromosomes (similar to a detailed road map) (**Figure 10.11**). A physical map is a representation of the physical distance, in nucleotides, between genes or genetic markers. Both genetic linkage maps and physical maps are required to build a complete picture of the genome. Having a complete map of the genome makes it easier for researchers to study individual genes. Human genome maps help researchers in their efforts to identify human disease-causing genes related to illnesses such as cancer, heart disease, and cystic fibrosis, to name a few. In addition, genome mapping can be used to help identify organisms with beneficial traits, such as microbes with the ability to clean up pollutants or even prevent pollution. Research involving plant genome mapping may lead to methods that produce higher crop yields or to the development of plants that adapt better to climate change.

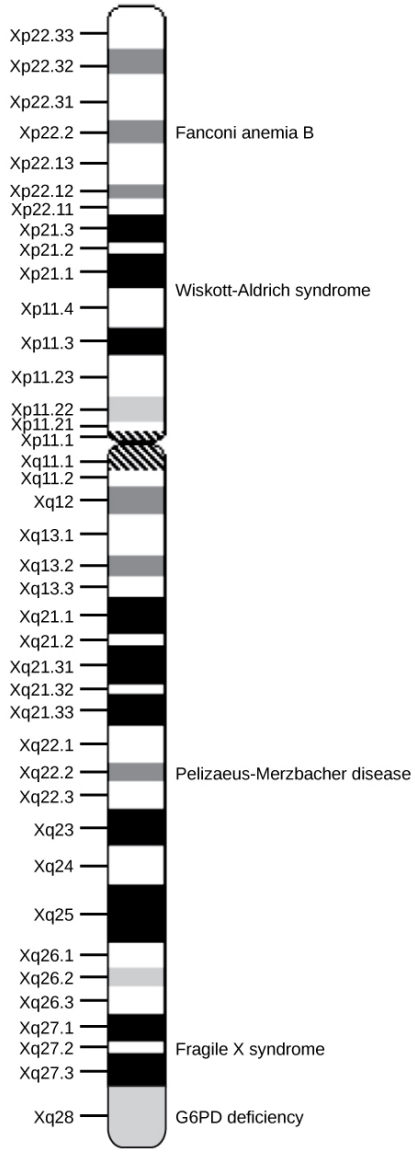


Figure 10.11 This is a physical map of the human X chromosome. (credit: modification of work by NCBI, NIH)

Genetic maps provide the outline, and physical maps provide the details. It is easy to understand why both types of genome-mapping techniques are important to show the big picture. Information obtained from each technique is used in combination to study the genome. Genomic mapping is used with different model organisms that are used for research. Genome mapping is still an ongoing process, and as more advanced techniques are developed, more advances are expected. Genome mapping is similar to completing a complicated puzzle using every piece of available data. Mapping information generated in laboratories all over the world is entered into central databases, such as the National Center for Biotechnology Information (NCBI). Efforts are made to make the information more easily accessible to researchers and the general public. Just as we use global positioning systems instead of paper maps to navigate through roadways, NCBI allows us to use a genome viewer tool to simplify the data mining process.

Online Mendelian Inheritance in Man (OMIM) (http://openstaxcollege.org/l/OMIM2) is a searchable online catalog of human genes and genetic disorders. This website shows genome mapping, and also details the history and research of each trait and disorder. Click the link to search for traits (such as handedness) and genetic disorders (such as diabetes).

Whole Genome Sequencing

Although there have been significant advances in the medical sciences in recent years, doctors are still confounded by many diseases and researchers are using whole genome sequencing to get to the bottom of the problem. **Whole genome sequencing** is a process that determines the DNA sequence of an entire genome. Whole genome sequencing is a brute-force approach to problem solving when there is a genetic basis at the core of a disease. Several laboratories now provide services to sequence, analyze, and interpret entire genomes.

In 2010, whole genome sequencing was used to save a young boy whose intestines had multiple mysterious abscesses. The child had several colon operations with no relief. Finally, a whole genome sequence revealed a defect in a pathway that controls apoptosis (programmed cell death). A bone marrow transplant was used to overcome this genetic disorder, leading to a cure for the boy. He was the first person to be successfully diagnosed using whole genome sequencing.

The first genomes to be sequenced, such as those belonging to viruses, bacteria, and yeast, were smaller in terms of the number of nucleotides than the genomes of multicellular organisms. The genomes of other model organisms, such as the mouse (*Mus musculus*), the fruit fly (*Drosophila melanogaster*), and the nematode (*Caenorhabditis elegans*) are now known. A great deal of basic research is performed in **model organisms** because the information can be applied to other organisms. A model organism is a species that is studied as a model to understand the biological processes in other species that can be represented by the model organism. For example, fruit flies are able to metabolize alcohol like humans, so the genes affecting sensitivity to alcohol have been studied in fruit flies in an effort to understand the variation in sensitivity to alcohol in humans. Having entire genomes sequenced helps with the research efforts in these model organisms (**Figure 10.12**).

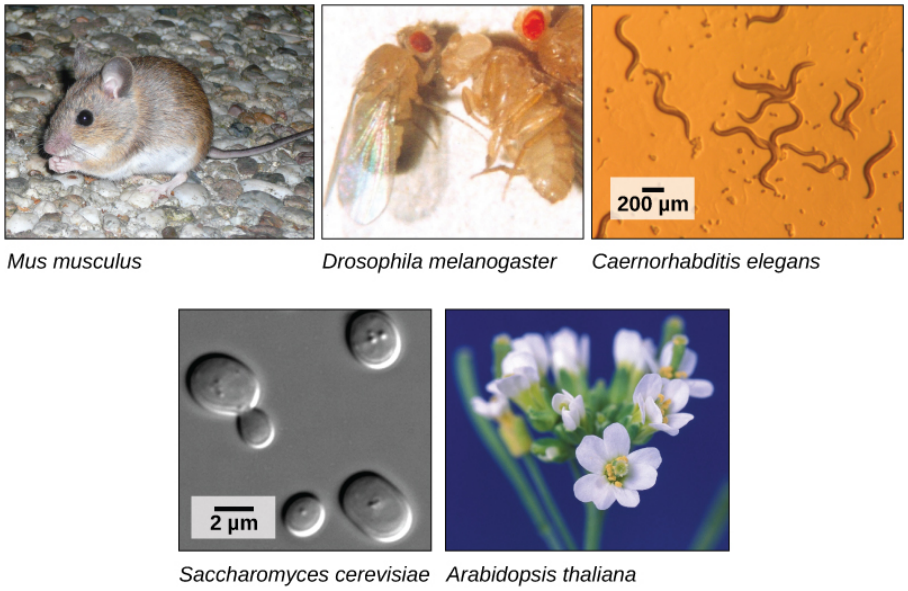


Figure 10.12 Much basic research is done with model organisms, such as the mouse, *Mus musculus*; the fruit fly, *Drosophila melanogaster*; the nematode *Caenorhabditis elegans*; the yeast *Saccharomyces cerevisiae*; and the common weed, *Arabidopsis thaliana*. (credit "mouse": modification of work by Florean Fortescue; credit "nematodes": modification of work by "snickclunk"/Flickr; credit "common weed": modification of work by Peggy Greb, USDA; scale-bar data from Matt Russell)

The first human genome sequence was published in 2003. The number of whole genomes that have been sequenced steadily increases and now includes hundreds of species and thousands of individual human genomes.

Applying Genomics

The introduction of DNA sequencing and whole genome sequencing projects, particularly the Human Genome Project, has expanded the applicability of DNA sequence information. Genomics is now being used in a wide variety of fields, such as metagenomics, pharmacogenomics, and mitochondrial genomics. The most commonly known application of genomics is to understand and find cures for diseases.

Predicting Disease Risk at the Individual Level

Predicting the risk of disease involves screening and identifying currently healthy individuals by genome analysis at the individual level. Intervention with lifestyle changes and drugs can be recommended before disease onset. However, this approach is most applicable when the problem arises from a single gene mutation. Such defects only account for about 5 percent of diseases found in developed countries. Most of the common diseases, such as heart disease, are multifactorial or polygenic, which refers to a phenotypic characteristic that is determined by two or more genes, and also environmental factors such as diet. In April 2010, scientists at Stanford University published the genome analysis of a healthy individual (Stephen Quake, a scientist at Stanford University, who had his genome sequenced); the analysis predicted his propensity to acquire various diseases. A risk assessment was done to analyze Quake's percentage of risk for 55 different medical conditions. A rare genetic mutation was found that showed him to be at risk for sudden heart attack. He was also predicted to have a 23 percent risk of developing prostate cancer and a 1.4 percent risk of developing Alzheimer's disease. The scientists used databases and several publications to analyze the genomic data. Even though genomic sequencing is becoming more affordable and analytical tools are becoming more reliable, ethical issues surrounding genomic analysis at a population level remain to be addressed. For example, could such data be legitimately used to charge more or less for insurance or to affect credit ratings?

Genome-wide Association Studies

Since 2005, it has been possible to conduct a type of study called a genome-wide association study, or GWAS. A GWAS is a method that identifies differences between individuals in single nucleotide polymorphisms (SNPs) that may be involved in causing diseases. The method is particularly suited to diseases that may be affected by one or many genetic changes throughout the genome. It is very difficult to identify the genes involved in such a disease using family history information. The GWAS method relies on a genetic database that has been in development since 2002 called the International HapMap Project. The HapMap Project sequenced the genomes of several hundred individuals from around the world and identified groups of SNPs. The groups include SNPs that are located near to each other on chromosomes so they tend to stay together

through recombination. The fact that the group stays together means that identifying one marker SNP is all that is needed to identify all the SNPs in the group. There are several million SNPs identified, but identifying them in other individuals who have not had their complete genome sequenced is much easier because only the marker SNPs need to be identified.

In a common design for a GWAS, two groups of individuals are chosen; one group has the disease, and the other group does not. The individuals in each group are matched in other characteristics to reduce the effect of confounding variables causing differences between the two groups. For example, the genotypes may differ because the two groups are mostly taken from different parts of the world. Once the individuals are chosen, and typically their numbers are a thousand or more for the study to work, samples of their DNA are obtained. The DNA is analyzed using automated systems to identify large differences in the percentage of particular SNPs between the two groups. Often the study examines a million or more SNPs in the DNA. The results of GWAS can be used in two ways: the genetic differences may be used as markers for susceptibility to the disease in undiagnosed individuals, and the particular genes identified can be targets for research into the molecular pathway of the disease and potential therapies. An offshoot of the discovery of gene associations with disease has been the formation of companies that provide so-called “personal genomics” that will identify risk levels for various diseases based on an individual’s SNP complement. The science behind these services is controversial.

Because GWAS looks for associations between genes and disease, these studies provide data for other research into causes, rather than answering specific questions themselves. An association between a gene difference and a disease does not necessarily mean there is a cause-and-effect relationship. However, some studies have provided useful information about the genetic causes of diseases. For example, three different studies in 2005 identified a gene for a protein involved in regulating inflammation in the body that is associated with a disease-causing blindness called age-related macular degeneration. This opened up new possibilities for research into the cause of this disease. A large number of genes have been identified to be associated with Crohn’s disease using GWAS, and some of these have suggested new hypothetical mechanisms for the cause of the disease.

Pharmacogenomics

Pharmacogenomics involves evaluating the effectiveness and safety of drugs on the basis of information from an individual's genomic sequence. Personal genome sequence information can be used to prescribe medications that will be most effective and least toxic on the basis of the individual patient’s genotype. Studying changes in gene expression could provide information about the gene transcription profile in the presence of the drug, which can be used as an early indicator of the potential for toxic effects. For example, genes involved in cellular growth and controlled cell death, when disturbed, could lead to the growth of cancerous cells. Genome-wide studies can also help to find new genes involved in drug toxicity. The gene signatures may not be completely accurate, but can be tested further before pathologic symptoms arise.

Metagenomics

Traditionally, microbiology has been taught with the view that microorganisms are best studied under pure culture conditions, which involves isolating a single type of cell and culturing it in the laboratory. Because microorganisms can go through several generations in a matter of hours, their gene expression profiles adapt to the new laboratory environment very quickly. On the other hand, many species resist being cultured in isolation. Most microorganisms do not live as isolated entities, but in microbial communities known as biofilms. For all of these reasons, pure culture is not always the best way to study microorganisms. **Metagenomics** is the study of the collective genomes of multiple species that grow and interact in an environmental niche. Metagenomics can be used to identify new species more rapidly and to analyze the effect of pollutants on the environment (**Figure 10.13**). Metagenomics techniques can now also be applied to communities of higher eukaryotes, such as fish.

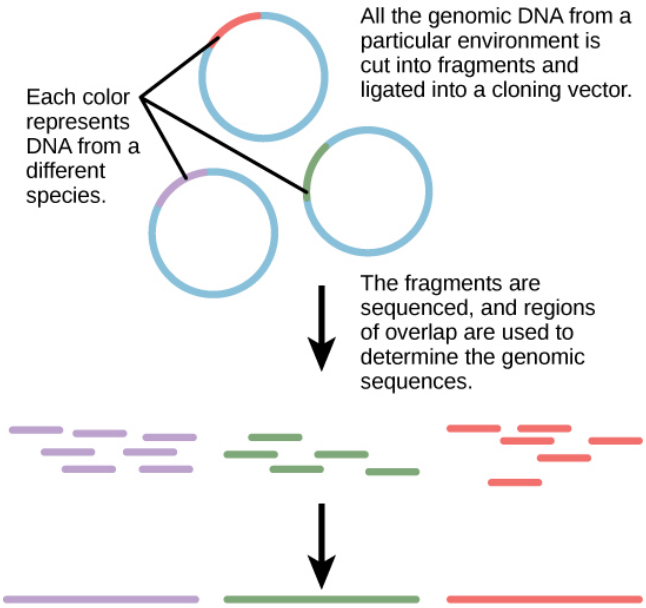


Figure 10.13 Metagenomics involves isolating DNA from multiple species within an environmental niche. The DNA is cut up and sequenced, allowing entire genome sequences of multiple species to be reconstructed from the sequences of overlapping pieces.

Creation of New Biofuels

Knowledge of the genomics of microorganisms is being used to find better ways to harness biofuels from algae and cyanobacteria. The primary sources of fuel today are coal, oil, wood, and other plant products such as ethanol. Although plants are renewable resources, there is still a need to find more alternative renewable sources of energy to meet our population's energy demands. The microbial world is one of the largest resources for genes that encode new enzymes and produce new organic compounds, and it remains largely untapped. This vast genetic resource holds the potential to provide new sources of biofuels (**Figure 10.14**).

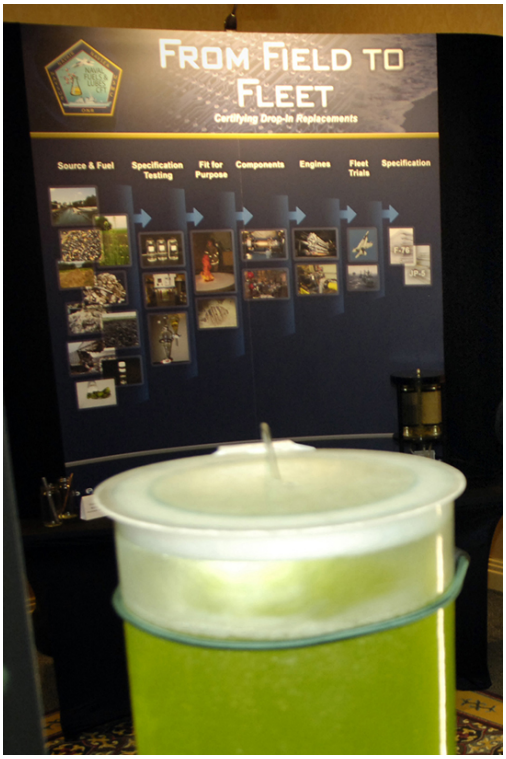


Figure 10.14 Renewable fuels were tested in Navy ships and aircraft at the first Naval Energy Forum. (credit: modification of work by John F. Williams, US Navy)

Mitochondrial Genomics

Mitochondria are intracellular organelles that contain their own DNA. Mitochondrial DNA mutates at a rapid rate and is often used to study evolutionary relationships. Another feature that makes studying the mitochondrial genome interesting is that in most multicellular organisms, the mitochondrial DNA is passed on from the mother during the process of fertilization. For this reason, mitochondrial genomics is often used to trace genealogy.

Genomics in Forensic Analysis

Information and clues obtained from DNA samples found at crime scenes have been used as evidence in court cases, and genetic markers have been used in forensic analysis. Genomic analysis has also become useful in this field. In 2001, the first use of genomics in forensics was published. It was a collaborative effort between academic research institutions and the FBI to solve the mysterious cases of anthrax (**Figure 10.15**) that was transported by the US Postal Service. Anthrax bacteria were made into an infectious powder and mailed to news media and two U.S. Senators. The powder infected the administrative staff and postal workers who opened or handled the letters. Five people died, and 17 were sickened from the bacteria. Using microbial genomics, researchers determined that a specific strain of anthrax was used in all the mailings; eventually, the source was traced to a scientist at a national biodefense laboratory in Maryland.

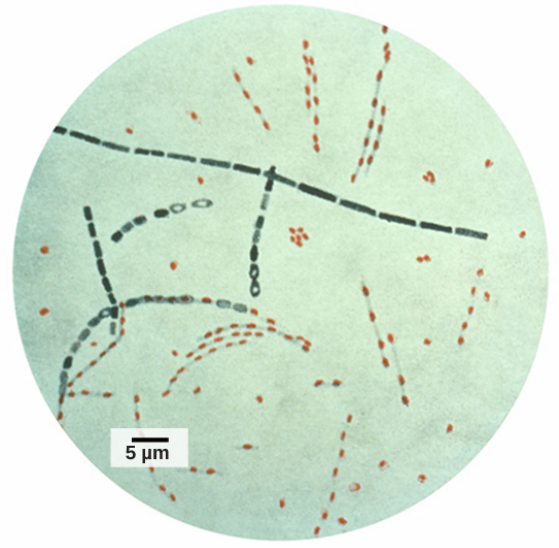


Figure 10.15 *Bacillus anthracis* is the organism that causes anthrax. (credit: modification of work by CDC; scale-bar data from Matt Russell)

Genomics in Agriculture

Genomics can reduce the trials and failures involved in scientific research to a certain extent, which could improve the quality and quantity of crop yields in agriculture (**Figure 10.16**). Linking traits to genes or gene signatures helps to improve crop breeding to generate hybrids with the most desirable qualities. Scientists use genomic data to identify desirable traits, and then transfer those traits to a different organism to create a new genetically modified organism, as described in the previous module. Scientists are discovering how genomics can improve the quality and quantity of agricultural production. For example, scientists could use desirable traits to create a useful product or enhance an existing product, such as making a drought-sensitive crop more tolerant of the dry season.

Figure 10.16 Transgenic agricultural plants can be made to resist disease. These transgenic plums are resistant to the plum pox virus. (credit: Scott Bauer, USDA ARS)

Proteomics

Proteins are the final products of genes that perform the function encoded by the gene. Proteins are composed of amino acids and play important roles in the cell. All enzymes (except ribozymes) are proteins and act as catalysts that affect the rate of reactions. Proteins are also regulatory molecules, and some are hormones. Transport proteins, such as hemoglobin, help transport oxygen to various organs. Antibodies that defend against foreign particles are also proteins. In the diseased state, protein function can be impaired because of changes at the genetic level or because of direct impact on a specific protein.

A proteome is the entire set of proteins produced by a cell type. Proteomes can be studied using the knowledge of genomes because genes code for mRNAs, and the mRNAs encode proteins. The study of the function of proteomes is called **proteomics**. Proteomics complements genomics and is useful when scientists want to test their hypotheses that were based on genes. Even though all cells in a multicellular organism have the same set of genes, the set of proteins produced in different tissues is different and dependent on gene expression. Thus, the genome is constant, but the proteome varies and is dynamic within an organism. In addition, RNAs can be alternatively spliced (cut and pasted to create novel combinations and novel proteins), and many proteins are modified after translation. Although the genome provides a blueprint, the final architecture depends on several factors that can change the progression of events that generate the proteome.

Genomes and proteomes of patients suffering from specific diseases are being studied to understand the genetic basis of the disease. The most prominent disease being studied with proteomic approaches is cancer (**Figure 10.17**). Proteomic approaches are being used to improve the screening and early detection of cancer; this is achieved by identifying proteins whose expression is affected by the disease process. An individual protein is called a **biomarker**, whereas a set of proteins with altered expression levels is called a **protein signature**. For a biomarker or protein signature to be useful as a candidate for early screening and detection of a cancer, it must be secreted in body fluids such as sweat, blood, or urine, so that large-scale screenings can be performed in a noninvasive fashion. The current problem with using biomarkers for the early detection of cancer is the high rate of false-negative results. A false-negative result is a negative test result that should have been positive. In other words, many cases of cancer go undetected, which makes biomarkers unreliable. Some examples of protein biomarkers used in cancer detection are CA-125 for ovarian cancer and PSA for prostate cancer. Protein signatures may be more reliable than biomarkers to detect cancer cells. Proteomics is also being used to develop individualized treatment plans, which involves the prediction of whether or not an individual will respond to specific drugs and the side effects that the individual may have. Proteomics is also being used to predict the possibility of disease recurrence.

Figure 10.17 This machine is preparing to do a proteomic pattern analysis to identify specific cancers so that an accurate cancer prognosis can be made. (credit: Dorie Hightower, NCI, NIH)

The National Cancer Institute has developed programs to improve the detection and treatment of cancer. The Clinical Proteomic Technologies for Cancer and the Early Detection Research Network are efforts to identify protein signatures specific to different types of cancers. The Biomedical Proteomics Program is designed to identify protein signatures and design effective therapies for cancer patients.

KEY TERMS

anneal in molecular biology, the process by which two single strands of DNA hydrogen bond at complementary nucleotides to form a double-stranded molecule

biomarker an individual protein that is uniquely produced in a diseased state

biotechnology the use of artificial methods to modify the genetic material of living organisms or cells to produce novel compounds or to perform new functions

cloning the production of an exact copy—specifically, an exact genetic copy—of a gene, cell, or organism

gel electrophoresis a technique used to separate molecules on the basis of their ability to migrate through a semisolid gel in response to an electric current

gene therapy the technique used to cure heritable diseases by replacing mutant genes with good genes

genetic engineering alteration of the genetic makeup of an organism using the molecular methods of biotechnology

genetic map an outline of genes and their location on a chromosome that is based on recombination frequencies between markers

genetic testing identifying gene variants in an individual that may lead to a genetic disease in that individual

genetically modified organism (GMO) an organism whose genome has been artificially changed

genomics the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species

metagenomics the study of the collective genomes of multiple species that grow and interact in an environmental niche

model organism a species that is studied and used as a model to understand the biological processes in other species represented by the model organism

pharmacogenomics the study of drug interactions with the genome or proteome; also called toxicogenomics

physical map a representation of the physical distance between genes or genetic markers

plasmid a small circular molecule of DNA found in bacteria that replicates independently of the main bacterial chromosome; plasmids code for some important traits for bacteria and can be used as vectors to transport DNA into bacteria in genetic engineering applications

polymerase chain reaction (PCR) a technique used to make multiple copies of DNA

protein signature a set of over- or under-expressed proteins characteristic of cells in a particular diseased tissue

proteomics study of the function of proteomes

recombinant DNA a combination of DNA fragments generated by molecular cloning that does not exist in nature

recombinant protein a protein that is expressed from recombinant DNA molecules

reproductive cloning cloning of entire organisms

restriction enzyme an enzyme that recognizes a specific nucleotide sequence in DNA and cuts the DNA double strand at that recognition site, often with a staggered cut leaving short single strands or “sticky” ends

reverse genetics a form of genetic analysis that manipulates DNA to disrupt or affect the product of a gene to analyze the gene’s function

transgenic describing an organism that receives DNA from a different species

whole genome sequencing a process that determines the nucleotide sequence of an entire genome

CHAPTER SUMMARY

10.1 Cloning and Genetic Engineering

Nucleic acids can be isolated from cells for the purposes of further analysis by breaking open the cells and enzymatically destroying all other major macromolecules. Fragmented or whole chromosomes can be separated on the basis of size by gel electrophoresis. Short stretches of DNA can be amplified by PCR. DNA can be cut (and subsequently re-spliced together) using restriction enzymes. The molecular and cellular techniques of biotechnology allow researchers to genetically engineer organisms, modifying them to achieve desirable traits.

Cloning may involve cloning small DNA fragments (molecular cloning), or cloning entire organisms (reproductive cloning). In molecular cloning with bacteria, a desired DNA fragment is inserted into a bacterial plasmid using restriction enzymes and the plasmid is taken up by a bacterium, which will then express the foreign DNA. Using other techniques, foreign genes can be inserted into eukaryotic organisms. In each case, the organisms are called transgenic organisms. In reproductive cloning, a donor nucleus is put into an enucleated egg cell, which is then stimulated to divide and develop into an organism.

In reverse genetics methods, a gene is mutated or removed in some way to identify its effect on the phenotype of the whole organism as a way to determine its function.

10.2 Biotechnology in Medicine and Agriculture

Genetic testing is performed to identify disease-causing genes, and can be used to benefit affected individuals and their relatives who have not developed disease symptoms yet. Gene therapy—by which functioning genes are incorporated into the genomes of individuals with a non-functioning mutant gene—has the potential to cure heritable diseases. Transgenic organisms possess DNA from a different species, usually generated by molecular cloning techniques. Vaccines, antibiotics, and hormones are examples of products obtained by recombinant DNA technology. Transgenic animals have been created for experimental purposes and some are used to produce some human proteins.

Genes are inserted into plants, using plasmids in the bacterium *Agrobacterium tumefaciens*, which infects plants. Transgenic plants have been created to improve the characteristics of crop plants—for example, by giving them insect resistance by inserting a gene for a bacterial toxin.

10.3 Genomics and Proteomics

Genome mapping is similar to solving a big, complicated puzzle with pieces of information coming from laboratories all over the world. Genetic maps provide an outline for the location of genes within a genome, and they estimate the distance between genes and genetic markers on the basis of the recombination frequency during meiosis. Physical maps provide detailed information about the physical distance between the genes. The most detailed information is available through sequence mapping. Information from all mapping and sequencing sources is combined to study an entire genome.

Whole genome sequencing is the latest available resource to treat genetic diseases. Some doctors are using whole genome sequencing to save lives. Genomics has many industrial applications, including biofuel development, agriculture, pharmaceuticals, and pollution control.

Imagination is the only barrier to the applicability of genomics. Genomics is being applied to most fields of biology; it can be used for personalized medicine, prediction of disease risks at an individual level, the study of drug interactions before the conduction of clinical trials, and the study of microorganisms in the environment as opposed to the laboratory. It is also being applied to the generation of new biofuels, genealogical assessment using mitochondria, advances in forensic science, and improvements in agriculture.

Proteomics is the study of the entire set of proteins expressed by a given type of cell under certain environmental conditions. In a multicellular organism, different cell types will have different proteomes, and these will vary with changes in the environment. Unlike a genome, a proteome is dynamic and under constant flux, which makes it more complicated and more useful than the knowledge of genomes alone.

ART CONNECTION QUESTIONS

1. **Figure 10.7** Why was Dolly a Finn-Dorset and not a Scottish Blackface sheep?

REVIEW QUESTIONS

2. In gel electrophoresis of DNA, the different bands in the final gel form because the DNA molecules _____.

a. are from different organisms
b. have different lengths
c. have different nucleotide compositions
d. have different genes

3. In the reproductive cloning of an animal, the genome of the cloned individual comes from _____.

a. a sperm cell
b. an egg cell
c. any gamete cell
d. a body cell

4. What carries a gene from one organism into a bacteria cell?

a. a plasmid
b. an electrophoresis gel
c. a restriction enzyme
d. polymerase chain reaction

5. What is a genetically modified organism (GMO)?

a. a plant with certain genes removed
b. an organism with an artificially altered genome
c. a hybrid organism
d. any agricultural organism produced by breeding or biotechnology

6. What is the role of *Agrobacterium tumefaciens* in the production of transgenic plants?

a. Genes from *A. tumefaciens* are inserted into plant DNA to give the plant different traits.
b. Transgenic plants have been given resistance to the pest *A. tumefaciens*.
c. *A. tumefaciens* is used as a vector to move genes into plant cells.
d. Plant genes are incorporated into the genome of *Agrobacterium tumefaciens*.

7. What is the most challenging issue facing genome sequencing?

a. the inability to develop fast and accurate sequencing techniques
b. the ethics of using information from genomes at the individual level
c. the availability and stability of DNA
d. all of the above

8. Genomics can be used in agriculture to:

a. generate new hybrid strains
b. improve disease resistance
c. improve yield
d. all of the above

9. What kind of diseases are studied using genome-wide association studies?

a. viral diseases
b. single-gene inherited diseases
c. diseases caused by multiple genes
d. diseases caused by environmental factors

CRITICAL THINKING QUESTIONS

10. What is the purpose and benefit of the polymerase chain reaction?

11. Today, it is possible for a diabetic patient to purchase human insulin from a pharmacist. What technology makes this possible and why is it a benefit over how things used to be?

12. Describe two of the applications for genome mapping.

13. Identify a possible advantage and a possible disadvantage of a genetic test that would identify genes in individuals that increase their probability of having Alzheimer's disease later in life.

11 | EVOLUTION AND ITS PROCESSES

Figure 11.1 The diversity of life on Earth is the result of evolution, a continuous process that is still occurring. (credit "wolf": modification of work by Gary Kramer, USFWS; credit "coral": modification of work by William Harrigan, NOAA; credit "river": modification of work by Vojtěch Dostál; credit "protozoa": modification of work by Sharon Franklin, Stephen Ausmus, USDA ARS; credit "fish" modification of work by Christian Mehlführer; credit "mushroom", "bee": modification of work by Cory Zanker; credit "tree": modification of work by Joseph Kranak)

Chapter Outline

Introduction

All species of living organisms—from the bacteria on our skin, to the trees in our yards, to the birds outside—evolved at some point from a different species. Although it may seem that living things today stay much the same from generation to generation, that is not the case: evolution is ongoing. Evolution is the process through which the characteristics of species change and through which new species arise.

The theory of evolution is the unifying theory of biology, meaning it is the framework within which biologists ask questions about the living world. Its power is that it provides direction for predictions about living things that are borne out in experiment after experiment. The Ukrainian-born American geneticist Theodosius Dobzhansky famously wrote that "nothing makes sense in biology except in the light of evolution."[1] He meant that the principle that all life has evolved

1. Theodosius Dobzhansky. "Biology, Molecular and Organismic." *American Zoologist* 4, no. 4 (1964): 449.

and diversified from a common ancestor is the foundation from which we understand all other questions in biology. This chapter will explain some of the mechanisms for evolutionary change and the kinds of questions that biologists can and have answered using evolutionary theory.

11.1 | Discovering How Populations Change

By the end of this section, you will be able to:

- Explain how Darwin's theory of evolution differed from the current view at the time
- Describe how the present-day theory of evolution was developed
- Describe how population genetics is used to study the evolution of populations

The theory of evolution by natural selection describes a mechanism for species change over time. That species change had been suggested and debated well before Darwin. The view that species were static and unchanging was grounded in the writings of Plato, yet there were also ancient Greeks that expressed evolutionary ideas.

In the eighteenth century, ideas about the evolution of animals were reintroduced by the naturalist Georges-Louis Leclerc, Comte de Buffon and even by Charles Darwin's grandfather, Erasmus Darwin. During this time, it was also accepted that there were extinct species. At the same time, James Hutton, the Scottish naturalist, proposed that geological change occurred gradually by the accumulation of small changes from processes (over long periods of time) just like those happening today. This contrasted with the predominant view that the geology of the planet was a consequence of catastrophic events occurring during a relatively brief past. Hutton's view was later popularized by the geologist Charles Lyell in the nineteenth century. Lyell became a friend to Darwin and his ideas were very influential on Darwin's thinking. Lyell argued that the greater age of Earth gave more time for gradual change in species, and the process provided an analogy for gradual change in species.

In the early nineteenth century, Jean-Baptiste Lamarck published a book that detailed a mechanism for evolutionary change that is now referred to as **inheritance of acquired characteristics**. In Lamarck's theory, modifications in an individual caused by its environment, or the use or disuse of a structure during its lifetime, could be inherited by its offspring and, thus, bring about change in a species. While this mechanism for evolutionary change as described by Lamarck was discredited, Lamarck's ideas were an important influence on evolutionary thought. The inscription on the statue of Lamarck that stands at the gates of the Jardin des Plantes in Paris describes him as the "founder of the doctrine of evolution."

Charles Darwin and Natural Selection

The actual mechanism for evolution was independently conceived of and described by two naturalists, Charles Darwin and Alfred Russell Wallace, in the mid-nineteenth century. Importantly, each spent time exploring the natural world on expeditions to the tropics. From 1831 to 1836, Darwin traveled around the world on *H.M.S. Beagle*, visiting South America, Australia, and the southern tip of Africa. Wallace traveled to Brazil to collect insects in the Amazon rainforest from 1848 to 1852 and to the Malay Archipelago from 1854 to 1862. Darwin's journey, like Wallace's later journeys in the Malay Archipelago, included stops at several island chains, the last being the Galápagos Islands (west of Ecuador). On these islands, Darwin observed species of organisms on different islands that were clearly similar, yet had distinct differences. For example, the ground finches inhabiting the Galápagos Islands comprised several species that each had a unique beak shape (**Figure 11.2**). He observed both that these finches closely resembled another finch species on the mainland of South America and that the group of species in the Galápagos formed a graded series of beak sizes and shapes, with very small differences between the most similar. Darwin imagined that the island species might be all species modified from one original mainland species. In 1860, he wrote, "Seeing this gradation and diversity of structure in one small, intimately related group of birds, one might really fancy that from an original paucity of birds in this archipelago, one species had been taken and modified for different ends."[2]

2. Charles Darwin, *Journal of Researches into the Natural History and Geology of the Countries Visited during the Voyage of H.M.S. Beagle Round the World, under the Command of Capt. Fitz Roy, R.N*, 2nd. ed. (London: John Murray, 1860), http://www.archive.org/details/journalofresea00darw.

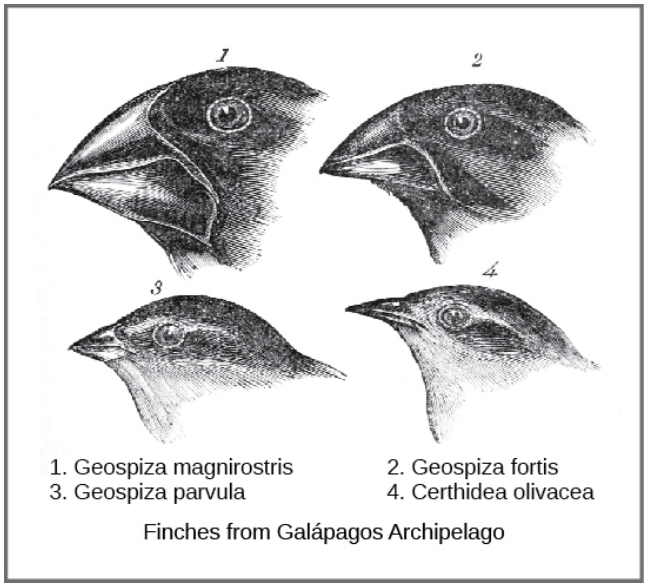


Figure 11.2 Darwin observed that beak shape varies among finch species. He postulated that the beak of an ancestral species had adapted over time to equip the finches to acquire different food sources. This illustration shows the beak shapes for four species of ground finch: 1. *Geospiza magnirostris* (the large ground finch), 2. *G. fortis* (the medium ground finch), 3. *G. parvula* (the small tree finch), and 4. *Certhidea olivacea* (the green-warbler finch).

Wallace and Darwin both observed similar patterns in other organisms and independently conceived a mechanism to explain how and why such changes could take place. Darwin called this mechanism natural selection. **Natural selection**, Darwin argued, was an inevitable outcome of three principles that operated in nature. First, the characteristics of organisms are inherited, or passed from parent to offspring. Second, more offspring are produced than are able to survive; in other words, resources for survival and reproduction are limited. The capacity for reproduction in all organisms outstrips the availability of resources to support their numbers. Thus, there is a competition for those resources in each generation. Both Darwin and Wallace's understanding of this principle came from reading an essay by the economist Thomas Malthus, who discussed this principle in relation to human populations. Third, offspring vary among each other in regard to their characteristics and those variations are inherited. Out of these three principles, Darwin and Wallace reasoned that offspring with inherited characteristics that allow them to best compete for limited resources will survive and have more offspring than those individuals with variations that are less able to compete. Because characteristics are inherited, these traits will be better represented in the next generation. This will lead to change in populations over generations in a process that Darwin called "descent with modification."

Papers by Darwin and Wallace (**Figure 11.3**) presenting the idea of natural selection were read together in 1858 before the Linnaean Society in London. The following year Darwin's book, *On the Origin of Species*, was published, which outlined in considerable detail his arguments for evolution by natural selection.

(a) (b)

Figure 11.3 (a) Charles Darwin and (b) Alfred Wallace wrote scientific papers on natural selection that were presented together before the Linnean Society in 1858.

Demonstrations of evolution by natural selection can be time consuming. One of the best demonstrations has been in the very birds that helped to inspire the theory, the Galápagos finches. Peter and Rosemary Grant and their colleagues have studied Galápagos finch populations every year since 1976 and have provided important demonstrations of the operation of natural selection. The Grants found changes from one generation to the next in the beak shapes of the medium ground finches on the Galápagos island of Daphne Major. The medium ground finch feeds on seeds. The birds have inherited variation in the bill shape with some individuals having wide, deep bills and others having thinner bills. Large-billed birds feed more efficiently on large, hard seeds, whereas smaller billed birds feed more efficiently on small, soft seeds. During 1977, a drought period altered vegetation on the island. After this period, the number of seeds declined dramatically: the decline in small, soft seeds was greater than the decline in large, hard seeds. The large-billed birds were able to survive better than the small-billed birds the following year. The year following the drought when the Grants measured beak sizes in the much-reduced population, they found that the average bill size was larger (**Figure 11.4**). This was clear evidence for natural selection (differences in survival) of bill size caused by the availability of seeds. The Grants had studied the inheritance of bill sizes and knew that the surviving large-billed birds would tend to produce offspring with larger bills, so the selection would lead to evolution of bill size. Subsequent studies by the Grants have demonstrated selection on and evolution of bill size in this species in response to changing conditions on the island. The evolution has occurred both to larger bills, as in this case, and to smaller bills when large seeds became rare.

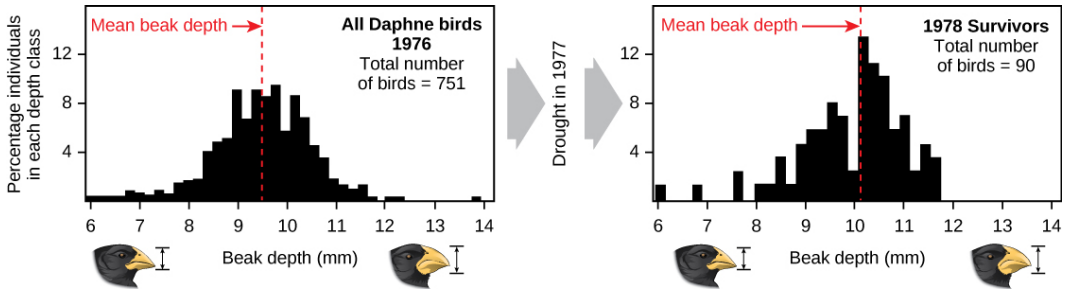


Figure 11.4 A drought on the Galápagos island of Daphne Major in 1977 reduced the number of small seeds available to finches, causing many of the small-beaked finches to die. This caused an increase in the finches' average beak size between 1976 and 1978.

Variation and Adaptation

Natural selection can only take place if there is **variation**, or differences, among individuals in a population. Importantly, these differences must have some genetic basis; otherwise, selection will not lead to change in the next generation. This is critical because variation among individuals can be caused by non-genetic reasons, such as an individual being taller because of better nutrition rather than different genes.

Genetic diversity in a population comes from two main sources: mutation and sexual reproduction. Mutation, a change in DNA, is the ultimate source of new alleles or new genetic variation in any population. An individual that has a mutated gene might have a different trait than other individuals in the population. However, this is not always the case. A mutation can have one of three outcomes on the organisms' appearance (or phenotype):

- A mutation may affect the phenotype of the organism in a way that gives it reduced fitness—lower likelihood of survival, resulting in fewer offspring.
- A mutation may produce a phenotype with a beneficial effect on fitness.
- Many mutations, called neutral mutations, will have no effect on fitness.

Mutations may also have a whole range of effect sizes on the fitness of the organism that expresses them in their phenotype, from a small effect to a great effect. Sexual reproduction and crossing over in meiosis also lead to genetic diversity: when two parents reproduce, unique combinations of alleles assemble to produce unique genotypes and, thus, phenotypes in each of the offspring.

A heritable trait that aids the survival and reproduction of an organism in its present environment is called an **adaptation**. An adaptation is a "match" of the organism to the environment. Adaptation to an environment comes about when a change in the range of genetic variation occurs over time that increases or maintains the match of the population with its environment. The variations in finch beaks shifted from generation to generation providing adaptation to food availability.

Whether or not a trait is favorable depends on the environment at the time. The same traits do not always have the same relative benefit or disadvantage because environmental conditions can change. For example, finches with large bills were benefited in one climate, while small bills were a disadvantage; in a different climate, the relationship reversed.

Patterns of Evolution

The evolution of species has resulted in enormous variation in form and function. When two species evolve in different directions from a common point, it is called **divergent evolution**. Such divergent evolution can be seen in the forms of the reproductive organs of flowering plants, which share the same basic anatomies; however, they can look very different as a result of selection in different physical environments, and adaptation to different kinds of pollinators (Figure 11.5).

(a) (b)

Figure 11.5 Flowering plants evolved from a common ancestor. Notice that the (a) dense blazing star and (b) purple coneflower vary in appearance, yet both share a similar basic morphology. (credit a, b: modification of work by Cory Zanker)

In other cases, similar phenotypes evolve independently in distantly related species. For example, flight has evolved in both bats and insects, and they both have structures we refer to as wings, which are adaptations to flight. The wings of bats and insects, however, evolved from very different original structures. When similar structures arise through evolution independently in different species it is called **convergent evolution**. The wings of bats and insects are called **analogous structures**; they are similar in function and appearance, but do not share an origin in a common ancestor. Instead they evolved independently in the two lineages. The wings of a hummingbird and an ostrich are **homologous structures**, meaning they share similarities (despite their differences resulting from evolutionary divergence). The wings of hummingbirds and ostriches did not evolve independently in the hummingbird lineage and the ostrich lineage—they descended from a common ancestor with wings.

The Modern Synthesis

The mechanisms of inheritance, genetics, were not understood at the time Darwin and Wallace were developing their idea of natural selection. This lack of understanding was a stumbling block to comprehending many aspects of evolution. In fact, blending inheritance was the predominant (and incorrect) genetic theory of the time, which made it difficult to understand how natural selection might operate. Darwin and Wallace were unaware of the genetics work by Austrian monk Gregor Mendel, which was published in 1866, not long after publication of *On the Origin of Species*. Mendel's work was rediscovered in the early twentieth century at which time geneticists were rapidly coming to an understanding of the basics of inheritance. Initially, the newly discovered particulate nature of genes made it difficult for biologists to understand how gradual evolution could occur. But over the next few decades genetics and evolution were integrated in what became known

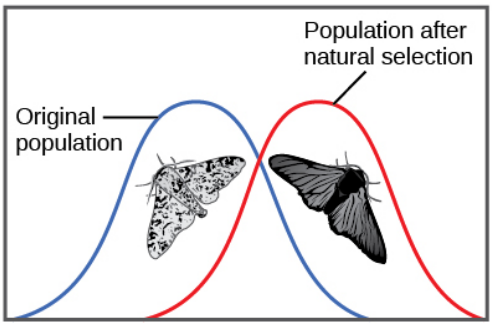
as the **modern synthesis**—the coherent understanding of the relationship between natural selection and genetics that took shape by the 1940s and is generally accepted today. In sum, the modern synthesis describes how evolutionary pressures, such as natural selection, can affect a population's genetic makeup, and, in turn, how this can result in the gradual evolution of populations and species. The theory also connects this gradual change of a population over time, called **microevolution**, with the processes that gave rise to new species and higher taxonomic groups with widely divergent characters, called **macroevolution**.

Population Genetics

Recall that a gene for a particular character may have several variants, or alleles, that code for different traits associated with that character. For example, in the ABO blood type system in humans, three alleles determine the particular blood-type protein on the surface of red blood cells. Each individual in a population of diploid organisms can only carry two alleles for a particular gene, but more than two may be present in the individuals that make up the population. Mendel followed alleles as they were inherited from parent to offspring. In the early twentieth century, biologists began to study what happens to all the alleles in a population in a field of study known as **population genetics**.

Until now, we have defined evolution as a change in the characteristics of a population of organisms, but behind that phenotypic change is genetic change. In population genetic terms, evolution is defined as a change in the frequency of an allele in a population. Using the ABO system as an example, the frequency of one of the alleles, I^A, is the number of copies of that allele divided by all the copies of the ABO gene in the population. For example, a study in Jordan found a frequency of I^A to be 26.1 percent.[3] The I^B, I^0 alleles made up 13.4 percent and 60.5 percent of the alleles respectively, and all of the frequencies add up to 100 percent. A change in this frequency over time would constitute evolution in the population.

There are several ways the allele frequencies of a population can change. One of those ways is natural selection. If a given allele confers a phenotype that allows an individual to have more offspring that survive and reproduce, that allele, by virtue of being inherited by those offspring, will be in greater frequency in the next generation. Since allele frequencies always add up to 100 percent, an increase in the frequency of one allele always means a corresponding decrease in one or more of the other alleles. Highly beneficial alleles may, over a very few generations, become "fixed" in this way, meaning that every individual of the population will carry the allele. Similarly, detrimental alleles may be swiftly eliminated from the **gene pool**, the sum of all the alleles in a population. Part of the study of population genetics is tracking how selective forces change the allele frequencies in a population over time, which can give scientists clues regarding the selective forces that may be operating on a given population. The studies of changes in wing coloration in the peppered moth from mottled white to dark in response to soot-covered tree trunks and then back to mottled white when factories stopped producing so much soot is a classic example of studying evolution in natural populations (Figure 11.6).



Light-colored peppered moths are better camouflaged against a pristine environment; likewise, dark-colored peppered moths are better camouflaged against a sooty environment. Thus, as the Industrial Revolution progressed in nineteenth-century England, the color of the moth population shifted from light to dark.

Figure 11.6 As the Industrial Revolution caused trees to darken from soot, darker colored peppered moths were better camouflaged than the lighter colored ones, which caused there to be more of the darker colored moths in the population.

In the early twentieth century, English mathematician Godfrey Hardy and German physician Wilhelm Weinberg independently provided an explanation for a somewhat counterintuitive concept. Hardy's original explanation was in response to a misunderstanding as to why a "dominant" allele, one that masks a recessive allele, should not increase in frequency in a population until it eliminated all the other alleles. The question resulted from a common confusion about what "dominant" means, but it forced Hardy, who was not even a biologist, to point out that if there are no factors that affect an allele frequency those frequencies will remain constant from one generation to the next. This principle is now known as the Hardy-Weinberg equilibrium. The theory states that a population's allele and genotype frequencies are inherently stable—unless some kind of evolutionary force is acting on the population, the population would carry the same alleles in the same proportions generation after generation. Individuals would, as a whole, look essentially the same and this would

3. Sahar S. Hanania, Dhia S. Hassawi, and Nidal M. Irshaid, "Allele Frequency and Molecular Genotypes of ABO Blood Group System in a Jordanian Population," *Journal of Medical Sciences* 7 (2007): 51-58, doi:10.3923/jms.2007.51.58

be unrelated to whether the alleles were dominant or recessive. The four most important evolutionary forces, which will disrupt the equilibrium, are natural selection, mutation, **genetic drift**, and **migration** into or out of a population. A fifth factor, nonrandom mating, will also disrupt the Hardy-Weinberg equilibrium but only by shifting genotype frequencies, not allele frequencies. In nonrandom mating, individuals are more likely to mate with like individuals (or unlike individuals) rather than at random. Since nonrandom mating does not change allele frequencies, it does not cause evolution directly. Natural selection has been described. Mutation creates one allele out of another one and changes an allele's frequency by a small, but continuous amount each generation. Each allele is generated by a low, constant mutation rate that will slowly increase the allele's frequency in a population if no other forces act on the allele. If natural selection acts against the allele, it will be removed from the population at a low rate leading to a frequency that results from a balance between selection and mutation. This is one reason that genetic diseases remain in the human population at very low frequencies. If the allele is favored by selection, it will increase in frequency. Genetic drift causes random changes in allele frequencies when populations are small. Genetic drift can often be important in evolution, as discussed in the next section. Finally, if two populations of a species have different allele frequencies, migration of individuals between them will cause frequency changes in both populations. As it happens, there is no population in which one or more of these processes are not operating, so populations are always evolving, and the Hardy-Weinberg equilibrium will never be exactly observed. However, the Hardy-Weinberg principle gives scientists a baseline expectation for allele frequencies in a non-evolving population to which they can compare evolving populations and thereby infer what evolutionary forces might be at play. The population is evolving if the frequencies of alleles or genotypes deviate from the value expected from the Hardy-Weinberg principle.

Darwin identified a special case of natural selection that he called sexual selection. Sexual selection affects an individual's ability to mate and thus produce offspring, and it leads to the evolution of dramatic traits that often appear maladaptive in terms of survival but persist because they give their owners greater reproductive success. Sexual selection occurs in two ways: through male–male competition for mates and through female selection of mates. Male–male competition takes the form of conflicts between males, which are often ritualized, but may also pose significant threats to a male's survival. Sometimes the competition is for territory, with females more likely to mate with males with higher quality territories. Female choice occurs when females choose a male based on a particular trait, such as feather colors, the performance of a mating dance, or the building of an elaborate structure. In some cases male–male competition and female choice combine in the mating process. In each of these cases, the traits selected for, such as fighting ability or feather color and length, become enhanced in the males. In general, it is thought that sexual selection can proceed to a point at which natural selection against a character's further enhancement prevents its further evolution because it negatively impacts the male's ability to survive. For example, colorful feathers or an elaborate display make the male more obvious to predators.

11.2 | Mechanisms of Evolution

By the end of this section, you will be able to:

- Describe the four basic causes of evolution: natural selection, mutation, genetic drift, and gene flow
- Explain how each evolutionary force can influence the allele frequencies of a population

The Hardy-Weinberg equilibrium principle says that allele frequencies in a population will remain constant in the absence of the four factors that could change them. Those factors are natural selection, mutation, genetic drift, and migration (gene flow). In fact, we know they are probably always affecting populations.

Natural Selection

Natural selection has already been discussed. Alleles are expressed in a phenotype. Depending on the environmental conditions, the phenotype confers an advantage or disadvantage to the individual with the phenotype relative to the other phenotypes in the population. If it is an advantage, then that individual will likely have more offspring than individuals with the other phenotypes, and this will mean that the allele behind the phenotype will have greater representation in the next generation. If conditions remain the same, those offspring, which are carrying the same allele, will also benefit. Over time, the allele will increase in frequency in the population.

Mutation

Mutation is a source of new alleles in a population. Mutation is a change in the DNA sequence of the gene. A mutation can change one allele into another, but the net effect is a change in frequency. The change in frequency resulting from mutation is small, so its effect on evolution is small unless it interacts with one of the other factors, such as selection. A mutation may produce an allele that is selected against, selected for, or selectively neutral. Harmful mutations are removed

from the population by selection and will generally only be found in very low frequencies equal to the mutation rate. Beneficial mutations will spread through the population through selection, although that initial spread is slow. Whether or not a mutation is beneficial or harmful is determined by whether it helps an organism survive to sexual maturity and reproduce. It should be noted that mutation is the ultimate source of genetic variation in all populations—new alleles, and, therefore, new genetic variations arise through mutation.

Genetic Drift

Another way a population's allele frequencies can change is genetic drift (**Figure 11.7**), which is simply the effect of chance. Genetic drift is most important in small populations. Drift would be completely absent in a population with infinite individuals, but, of course, no population is this large. Genetic drift occurs because the alleles in an offspring generation are a random sample of the alleles in the parent generation. Alleles may or may not make it into the next generation due to chance events including mortality of an individual, events affecting finding a mate, and even the events affecting which gametes end up in fertilizations. If one individual in a population of ten individuals happens to die before it leaves any offspring to the next generation, all of its genes—a tenth of the population's gene pool—will be suddenly lost. In a population of 100, that 1 individual represents only 1 percent of the overall gene pool; therefore, it has much less impact on the population's genetic structure and is unlikely to remove all copies of even a relatively rare allele.

Imagine a population of ten individuals, half with allele *A* and half with allele *a* (the individuals are haploid). In a stable population, the next generation will also have ten individuals. Choose that generation randomly by flipping a coin ten times and let heads be *A* and tails be *a*. It is unlikely that the next generation will have exactly half of each allele. There might be six of one and four of the other, or some different set of frequencies. Thus, the allele frequencies have changed and evolution has occurred. A coin will no longer work to choose the next generation (because the odds are no longer one half for each allele). The frequency in each generation will drift up and down on what is known as a random walk until at one point either all *A* or all *a* are chosen and that allele is fixed from that point on. This could take a very long time for a large population. This simplification is not very biological, but it can be shown that real populations behave this way. The effect of drift on frequencies is greater the smaller a population is. Its effect is also greater on an allele with a frequency far from one half. Drift will influence every allele, even those that are being naturally selected.

art CONNECTION

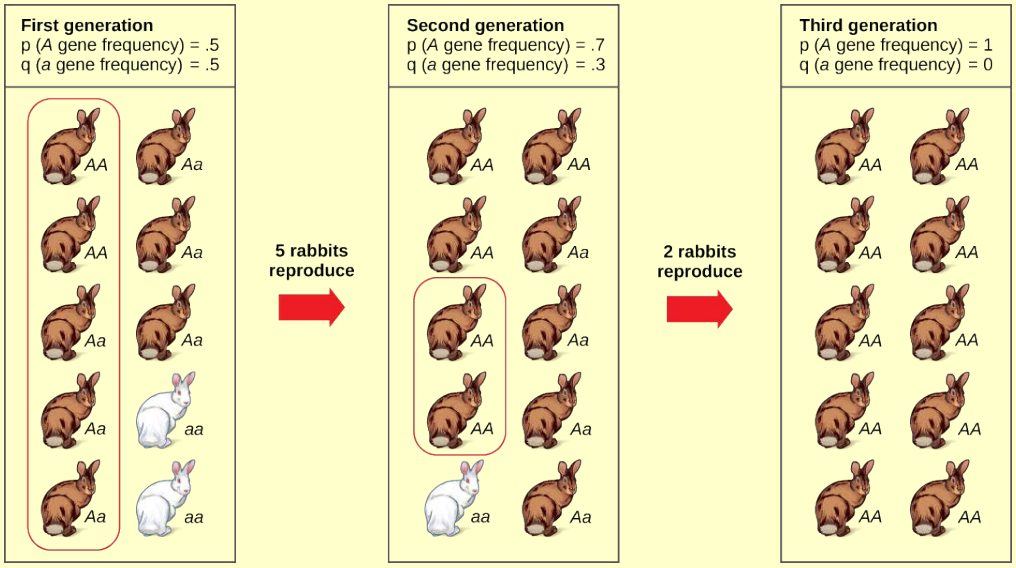


Figure 11.7 Genetic drift in a population can lead to the elimination of an allele from a population by chance. In each generation, a random set of individuals reproduces to produce the next generation. The frequency of alleles in the next generation is equal to the frequency of alleles among the individuals reproducing.

Do you think genetic drift would happen more quickly on an island or on the mainland?

Genetic drift can also be magnified by natural or human-caused events, such as a disaster that randomly kills a large portion of the population, which is known as the **bottleneck effect** that results in a large portion of the genome suddenly being wiped out (**Figure 11.8**). In one fell swoop, the genetic structure of the survivors becomes the genetic structure of the entire

population, which may be very different from the pre-disaster population. The disaster must be one that kills for reasons unrelated to the organism's traits, such as a hurricane or lava flow. A mass killing caused by unusually cold temperatures at night, is likely to affect individuals differently depending on the alleles they possess that confer cold hardiness.

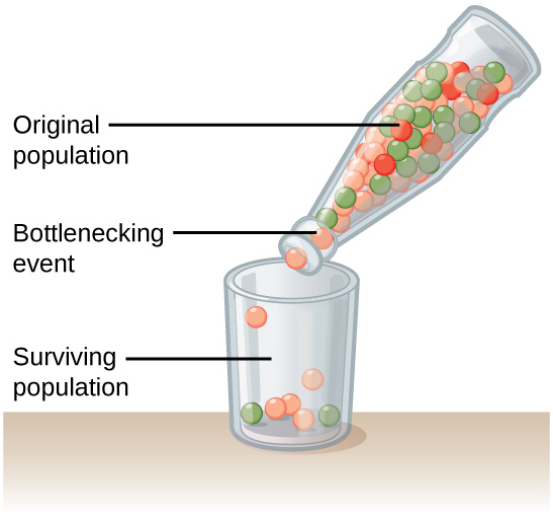


Figure 11.8 A chance event or catastrophe can reduce the genetic variability within a population.

Another scenario in which populations might experience a strong influence of genetic drift is if some portion of the population leaves to start a new population in a new location, or if a population gets divided by a physical barrier of some kind. In this situation, those individuals are unlikely to be representative of the entire population which results in the **founder effect**. The founder effect occurs when the genetic structure matches that of the new population's founding fathers and mothers. The founder effect is believed to have been a key factor in the genetic history of the Afrikaner population of Dutch settlers in South Africa, as evidenced by mutations that are common in Afrikaners but rare in most other populations. This is likely due to a higher-than-normal proportion of the founding colonists, which were a small sample of the original population, carried these mutations. As a result, the population expresses unusually high incidences of Huntington's disease (HD) and Fanconi anemia (FA), a genetic disorder known to cause bone marrow and congenital abnormalities, and even cancer.[4]

CONCEPT in ACTION

Visit this **site (http://openstaxcollege.org/l/genetic_drift2)** to learn more about genetic drift and to run simulations of allele changes caused by drift.

Gene Flow

Another important evolutionary force is **gene flow**, or the flow of alleles in and out of a population resulting from the migration of individuals or gametes (**Figure 11.9**). While some populations are fairly stable, others experience more flux. Many plants, for example, send their seeds far and wide, by wind or in the guts of animals; these seeds may introduce alleles common in the source population to a new population in which they are rare.

4. A. J. Tipping et al., "Molecular and Genealogical Evidence for a Founder Effect in Fanconi Anemia Families of the Afrikaner Population of South Africa," *PNAS* 98, no. 10 (2001): 5734-5739, doi: 10.1073/pnas.091402398.

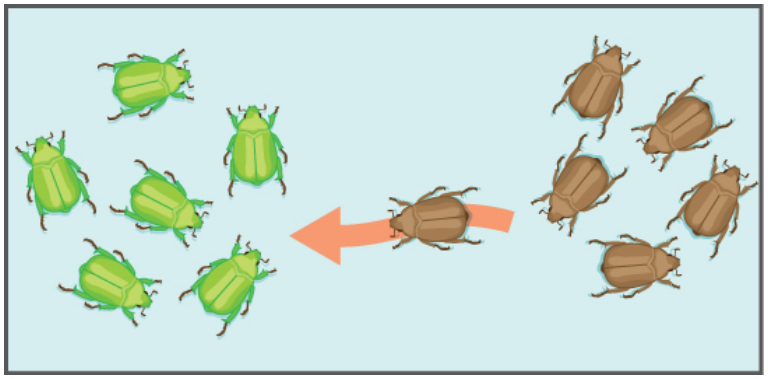


Figure 11.9 Gene flow can occur when an individual travels from one geographic location to another and joins a different population of the species. In the example shown here, the brown allele is introduced into the green population.

11.3 | Evidence of Evolution

By the end of this section, you will be able to:

- Explain sources of evidence for evolution
- Define homologous and vestigial structures

The evidence for evolution is compelling and extensive. Looking at every level of organization in living systems, biologists see the signature of past and present evolution. Darwin dedicated a large portion of his book, *On the Origin of Species*, identifying patterns in nature that were consistent with evolution and since Darwin our understanding has become clearer and broader.

Fossils

Fossils provide solid evidence that organisms from the past are not the same as those found today; fossils show a progression of evolution. Scientists determine the age of fossils and categorize them all over the world to determine when the organisms lived relative to each other. The resulting fossil record tells the story of the past, and shows the evolution of form over millions of years (**Figure 11.10**). For example, highly detailed fossil records have been recovered for sequences of species in the evolution of whales and modern horses. The fossil record of horses in North America is especially rich and many contain transition fossils: those showing intermediate anatomy between earlier and later forms. The fossil record extends back to a dog-like ancestor some 55 million years ago that gave rise to the first horse-like species 55 to 42 million years ago in the genus *Eohippus*. The series of fossils tracks the change in anatomy resulting from a gradual drying trend that changed the landscape from a forested one to a prairie. Successive fossils show the evolution of teeth shapes and foot and leg anatomy to a grazing habit, with adaptations for escaping predators, for example in species of *Mesohippus* found from 40 to 30 million years ago. Later species showed gains in size, such as those of *Hipparion*, which existed from about 23 to 2 million years ago. The fossil record shows several adaptive radiations in the horse lineage, which is now much reduced to only one genus, *Equus*, with several species.

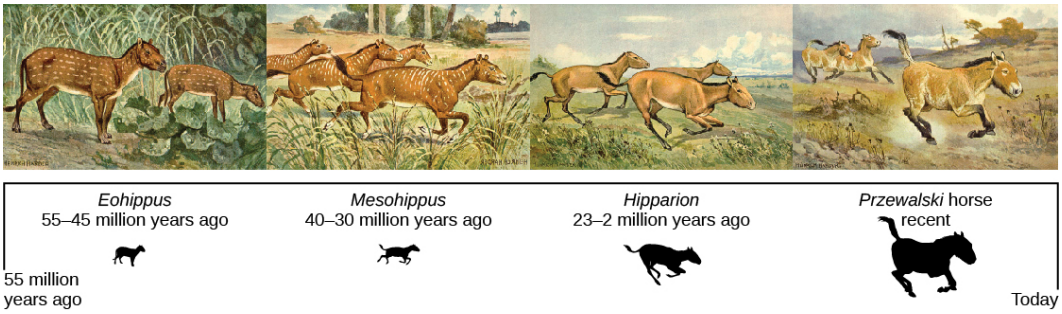


Figure 11.10 This illustration shows an artist's renderings of these species derived from fossils of the evolutionary history of the horse and its ancestors. The species depicted are only four from a very diverse lineage that contains many branches, dead ends, and adaptive radiations. One of the trends, depicted here is the evolutionary tracking of a drying climate and increase in prairie versus forest habitat reflected in forms that are more adapted to grazing and predator escape through running. Przewalski's horse is one of a few living species of horse.

Anatomy and Embryology

Another type of evidence for evolution is the presence of structures in organisms that share the same basic form. For example, the bones in the appendages of a human, dog, bird, and whale all share the same overall construction (**Figure 11.11**). That similarity results from their origin in the appendages of a common ancestor. Over time, evolution led to changes in the shapes and sizes of these bones in different species, but they have maintained the same overall layout, evidence of descent from a common ancestor. Scientists call these synonymous parts homologous structures. Some structures exist in organisms that have no apparent function at all, and appear to be residual parts from a past ancestor. For example, some snakes have pelvic bones despite having no legs because they descended from reptiles that did have legs. These unused structures without function are called **vestigial structures**. Other examples of vestigial structures are wings on flightless birds (which may have other functions), leaves on some cacti, traces of pelvic bones in whales, and the sightless eyes of cave animals.

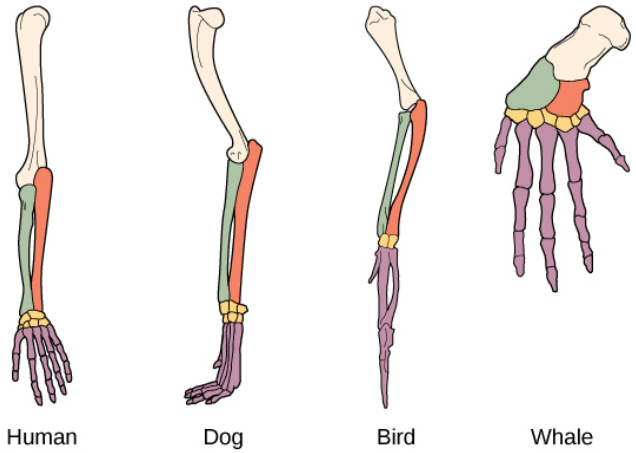


Figure 11.11 The similar construction of these appendages indicates that these organisms share a common ancestor.

CONCEPT in ACTION

openstax COLLEGE

Click through the activities at this **interactive site (http://openstaxcollege.org/l/bone_structure2)** to guess which bone structures are homologous and which are analogous, and to see examples of all kinds of evolutionary adaptations that illustrate these concepts.

Another evidence of evolution is the convergence of form in organisms that share similar environments. For example, species of unrelated animals, such as the arctic fox and ptarmigan (a bird), living in the arctic region have temporary white coverings during winter to blend with the snow and ice (**Figure 11.12**). The similarity occurs not because of common ancestry, indeed one covering is of fur and the other of feathers, but because of similar selection pressures—the benefits of not being seen by predators.

(a) (b)

Figure 11.12 The white winter coat of (a) the arctic fox and (b) the ptarmigan's plumage are adaptations to their environments. (credit a: modification of work by Keith Morehouse)

Embryology, the study of the development of the anatomy of an organism to its adult form also provides evidence of relatedness between now widely divergent groups of organisms. Structures that are absent in some groups often appear in their embryonic forms and disappear by the time the adult or juvenile form is reached. For example, all vertebrate embryos, including humans, exhibit gill slits at some point in their early development. These disappear in the adults of terrestrial groups, but are maintained in adult forms of aquatic groups such as fish and some amphibians. Great ape embryos, including humans, have a tail structure during their development that is lost by the time of birth. The reason embryos of unrelated species are often similar is that mutational changes that affect the organism during embryonic development can cause amplified differences in the adult, even while the embryonic similarities are preserved.

Biogeography

The geographic distribution of organisms on the planet follows patterns that are best explained by evolution in conjunction with the movement of tectonic plates over geological time. Broad groups that evolved before the breakup of the supercontinent Pangaea (about 200 million years ago) are distributed worldwide. Groups that evolved since the breakup appear uniquely in regions of the planet, for example the unique flora and fauna of northern continents that formed from the supercontinent Laurasia and of the southern continents that formed from the supercontinent Gondwana. The presence of Proteaceae in Australia, southern Africa, and South America is best explained by the plant family's presence there prior to the southern supercontinent Gondwana breaking up (**Figure 11.13**).

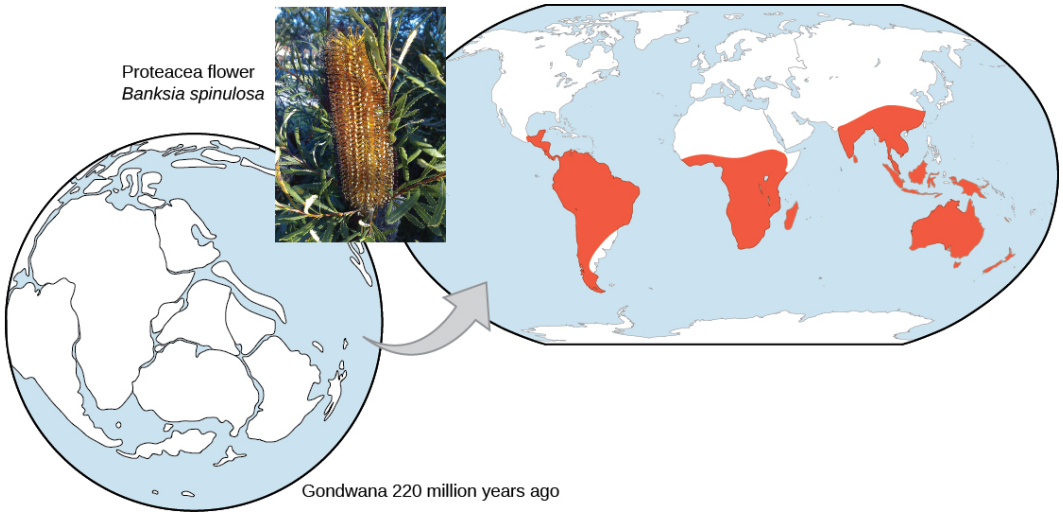


Figure 11.13 The Proteacea family of plants evolved before the supercontinent Gondwana broke up. Today, members of this plant family are found throughout the southern hemisphere (shown in red). (credit "Proteacea flower": modification of work by "dorofofoto"/Flickr)

The great diversification of the marsupials in Australia and the absence of other mammals reflects that island continent's long isolation. Australia has an abundance of endemic species—species found nowhere else—which is typical of islands whose isolation by expanses of water prevents migration of species to other regions. Over time, these species diverge evolutionarily into new species that look very different from their ancestors that may exist on the mainland. The marsupials of Australia, the finches on the Galápagos, and many species on the Hawaiian Islands are all found nowhere else but on their island, yet display distant relationships to ancestral species on mainlands.

Molecular Biology

Like anatomical structures, the structures of the molecules of life reflect descent with modification. Evidence of a common ancestor for all of life is reflected in the universality of DNA as the genetic material and of the near universality of the genetic code and the machinery of DNA replication and expression. Fundamental divisions in life between the three domains are reflected in major structural differences in otherwise conservative structures such as the components of ribosomes and the structures of membranes. In general, the relatedness of groups of organisms is reflected in the similarity of their DNA sequences—exactly the pattern that would be expected from descent and diversification from a common ancestor.

DNA sequences have also shed light on some of the mechanisms of evolution. For example, it is clear that the evolution of new functions for proteins commonly occurs after gene duplication events. These duplications are a kind of mutation in which an entire gene is added as an extra copy (or many copies) in the genome. These duplications allow the free modification of one copy by mutation, selection, and drift, while the second copy continues to produce a functional protein. This allows the original function for the protein to be kept, while evolutionary forces tweak the copy until it functions in a new way.

11.4 | Speciation

By the end of this section, you will be able to:

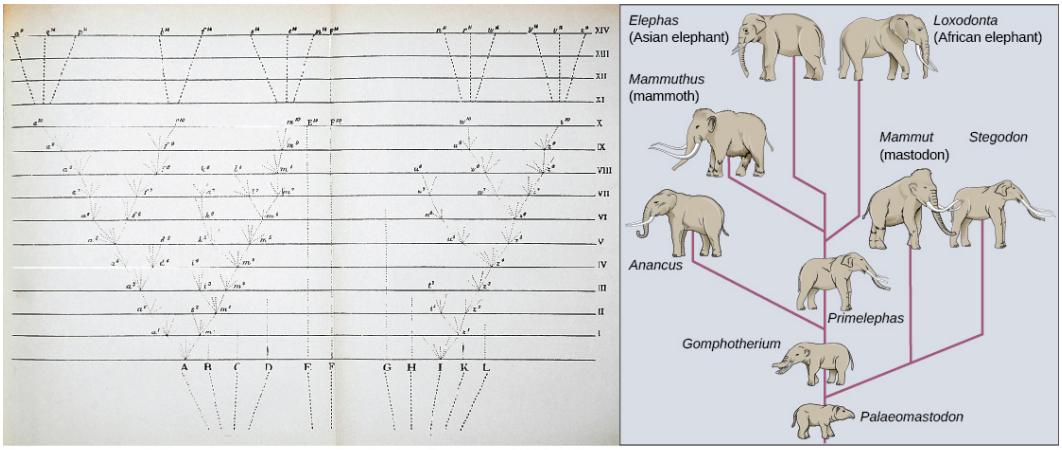
- Describe the definition of species and how species are identified as different
- Explain allopatric and sympatric speciation
- Describe adaptive radiation

The biological definition of species, which works for sexually reproducing organisms, is a group of actually or potentially interbreeding individuals. According to this definition, one species is distinguished from another by the possibility of matings between individuals from each species to produce fertile offspring. There are exceptions to this rule. Many species are similar enough that hybrid offspring are possible and may often occur in nature, but for the majority of species this

rule generally holds. In fact, the presence of hybrids between similar species suggests that they may have descended from a single interbreeding species and that the speciation process may not yet be completed.

Given the extraordinary diversity of life on the planet there must be mechanisms for **speciation**: the formation of two species from one original species. Darwin envisioned this process as a branching event and diagrammed the process in the only illustration found in *On the Origin of Species* (**Figure 11.14a**). For speciation to occur, two new populations must be formed from one original population, and they must evolve in such a way that it becomes impossible for individuals from the two new populations to interbreed. Biologists have proposed mechanisms by which this could occur that fall into two broad categories. **Allopatric speciation**, meaning speciation in "other homelands," involves a geographic separation of populations from a parent species and subsequent evolution. **Sympatric speciation**, meaning speciation in the "same homeland," involves speciation occurring within a parent species while remaining in one location.

Biologists think of speciation events as the splitting of one ancestral species into two descendant species. There is no reason why there might not be more than two species formed at one time except that it is less likely and such multiple events can also be conceptualized as single splits occurring close in time.



(a) (b)

Figure 11.14 The only illustration in Darwin's *On the Origin of Species* is (a) a diagram showing speciation events leading to biological diversity. The diagram shows similarities to phylogenetic charts that are drawn today to illustrate the relationships of species. (b) Modern elephants evolved from the *Palaeomastodon*, a species that lived in Egypt 35–50 million years ago.

Speciation through Geographic Separation

A geographically continuous population has a gene pool that is relatively homogeneous. Gene flow, the movement of alleles across the range of the species, is relatively free because individuals can move and then mate with individuals in their new location. Thus, the frequency of an allele at one end of a distribution will be similar to the frequency of the allele at the other end. When populations become geographically discontinuous that free-flow of alleles is prevented. When that separation lasts for a period of time, the two populations are able to evolve along different trajectories. Thus, their allele frequencies at numerous genetic loci gradually become more and more different as new alleles independently arise by mutation in each population. Typically, environmental conditions, such as climate, resources, predators, and competitors, for the two populations will differ causing natural selection to favor divergent adaptations in each group. Different histories of genetic drift, enhanced because the populations are smaller than the parent population, will also lead to divergence.

Given enough time, the genetic and phenotypic divergence between populations will likely affect characters that influence reproduction enough that were individuals of the two populations brought together, mating would be less likely, or if a mating occurred, offspring would be non-viable or infertile. Many types of diverging characters may affect the reproductive isolation (inability to interbreed) of the two populations. These mechanisms of reproductive isolation can be divided into prezygotic mechanisms (those that operate before fertilization) and postzygotic mechanisms (those that operate after fertilization). Prezygotic mechanisms include traits that allow the individuals to find each other, such as the timing of mating, sensitivity to pheromones, or choice of mating sites. If individuals are able to encounter each other, character divergence may prevent courtship rituals from leading to a mating either because female preferences have changed or male behaviors have changed. Physiological changes may interfere with successful fertilization if mating is able to occur. Postzygotic mechanisms include genetic incompatibilities that prevent proper development of the offspring, or if the offspring live, they may be unable to produce viable gametes themselves as in the example of the mule, the infertile offspring of a female horse and a male donkey.

If the two isolated populations are brought back together and the hybrid offspring that formed from matings between individuals of the two populations have lower survivorship or reduced fertility, then selection will favor individuals that are able to discriminate between potential mates of their own population and the other population. This selection will enhance the reproductive isolation.

Isolation of populations leading to allopatric speciation can occur in a variety of ways: from a river forming a new branch, erosion forming a new valley, or a group of organisms traveling to a new location without the ability to return, such as seeds floating over the ocean to an island. The nature of the geographic separation necessary to isolate populations depends entirely on the biology of the organism and its potential for dispersal. If two flying insect populations took up residence in separate nearby valleys, chances are that individuals from each population would fly back and forth, continuing gene flow. However, if two rodent populations became divided by the formation of a new lake, continued gene flow would be unlikely; therefore, speciation would be more likely.

Biologists group allopatric processes into two categories. If a few members of a species move to a new geographical area, this is called **dispersal**. If a natural situation arises to physically divide organisms, this is called **vicariance**.

Scientists have documented numerous cases of allopatric speciation taking place. For example, along the west coast of the United States, two separate subspecies of spotted owls exist. The northern spotted owl has genetic and phenotypic differences from its close relative, the Mexican spotted owl, which lives in the south (**Figure 11.15**). The cause of their initial separation is not clear, but it may have been caused by the glaciers of the ice age dividing an initial population into two.[5]

Figure 11.15 The northern spotted owl and the Mexican spotted owl inhabit geographically separate locations with different climates and ecosystems. The owl is an example of incipient speciation. (credit “northern spotted owl”: modification of work by John and Karen Hollingsworth, USFWS; credit “Mexican spotted owl”: modification of work by Bill Radke, USFWS)

Additionally, scientists have found that the further the distance between two groups that once were the same species, the more likely for speciation to take place. This seems logical because as the distance increases, the various environmental factors would likely have less in common than locations in close proximity. Consider the two owls; in the north, the climate is cooler than in the south; the other types of organisms in each ecosystem differ, as do their behaviors and habits; also, the hunting habits and prey choices of the owls in the south vary from the northern ones. These variances can lead to evolved differences in the owls, and over time speciation will likely occur unless gene flow between the populations is restored.

5. Courtney, S.P., et al, “Scientific Evaluation of the Status of the Northern Spotted Owl,” Sustainable Ecosystems Institute (2004), Portland, OR.

In some cases, a population of one species disperses throughout an area, and each finds a distinct niche or isolated habitat. Over time, the varied demands of their new lifestyles lead to multiple speciation events originating from a single species, which is called **adaptive radiation**. From one point of origin, many adaptations evolve causing the species to radiate into several new ones. Island archipelagos like the Hawaiian Islands provide an ideal context for adaptive radiation events because water surrounds each island, which leads to geographical isolation for many organisms (**Figure 11.16**). The Hawaiian honeycreeper illustrates one example of adaptive radiation. From a single species, called the founder species, numerous species have evolved, including the eight shown in **Figure 11.16**.

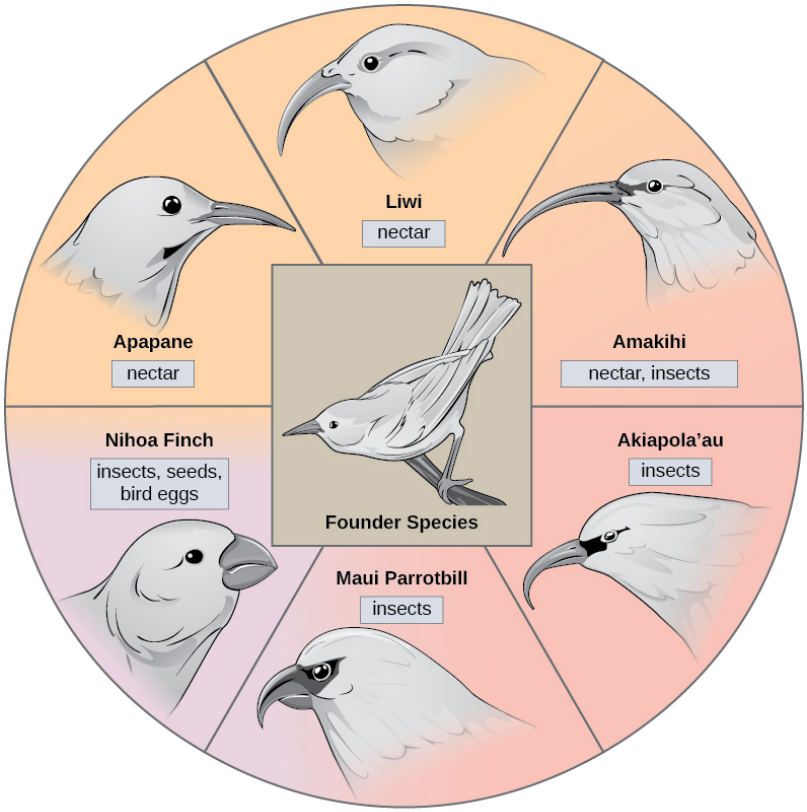


Figure 11.16 The honeycreeper birds illustrate adaptive radiation. From one original species of bird, multiple others evolved, each with its own distinctive characteristics.

Notice the differences in the species' beaks in **Figure 11.16**. Change in the genetic variation for beaks in response to natural selection based on specific food sources in each new habitat led to evolution of a different beak suited to the specific food source. The fruit and seed-eating birds have thicker, stronger beaks which are suited to break hard nuts. The nectar-eating birds have long beaks to dip into flowers to reach their nectar. The insect-eating birds have beaks like swords, appropriate for stabbing and impaling insects. Darwin's finches are another well-studied example of adaptive radiation in an archipelago.

CONCEPT in ACTION

Click through this **interactive site (http://openstaxcollege.org/l/bird_evolution)** to see how island birds evolved; click to see images of each species in evolutionary increments from five million years ago to today.

Speciation without Geographic Separation

Can divergence occur if no physical barriers are in place to separate individuals who continue to live and reproduce in the same habitat? A number of mechanisms for sympatric speciation have been proposed and studied.

One form of sympatric speciation can begin with a chromosomal error during meiosis or the formation of a hybrid individual with too many chromosomes. Polyploidy is a condition in which a cell, or organism, has an extra set, or sets, of chromosomes. Scientists have identified two main types of polyploidy that can lead to reproductive isolation of an individual in the polyploid state. In some cases a polyploid individual will have two or more complete sets of chromosomes from its own species in a condition called autopolyploidy (**Figure 11.17**). The prefix "auto" means self, so the term means multiple chromosomes from one's own species. Polyploidy results from an error in meiosis in which all of the chromosomes move into one cell instead of separating.

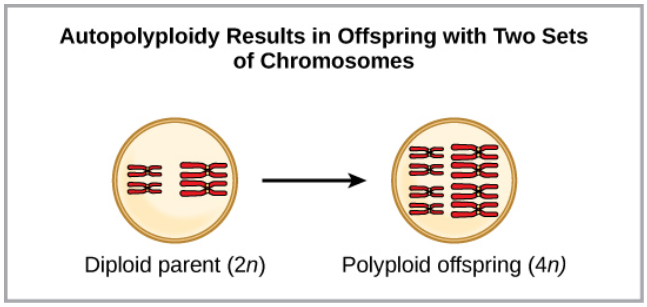


Figure 11.17 Autopolyploidy results when mitosis is not followed by cytokinesis.

For example, if a plant species with $2n = 6$ produces autopolyploid gametes that are also diploid ($2n = 6$, when they should be $n = 3$), the gametes now have twice as many chromosomes as they should have. These new gametes will be incompatible with the normal gametes produced by this plant species. But they could either self-pollinate or reproduce with other autopolyploid plants with gametes having the same diploid number. In this way, sympatric speciation can occur quickly by forming offspring with $4n$ called a tetraploid. These individuals would immediately be able to reproduce only with those of this new kind and not those of the ancestral species. The other form of polyploidy occurs when individuals of two different species reproduce to form a viable offspring called an allopolyploid. The prefix "allo" means "other" (recall from allopatric); therefore, an allopolyploid occurs when gametes from two different species combine. **Figure 11.18** illustrates one possible way an allopolyploidy can form. Notice how it takes two generations, or two reproductive acts, before the viable fertile hybrid results.

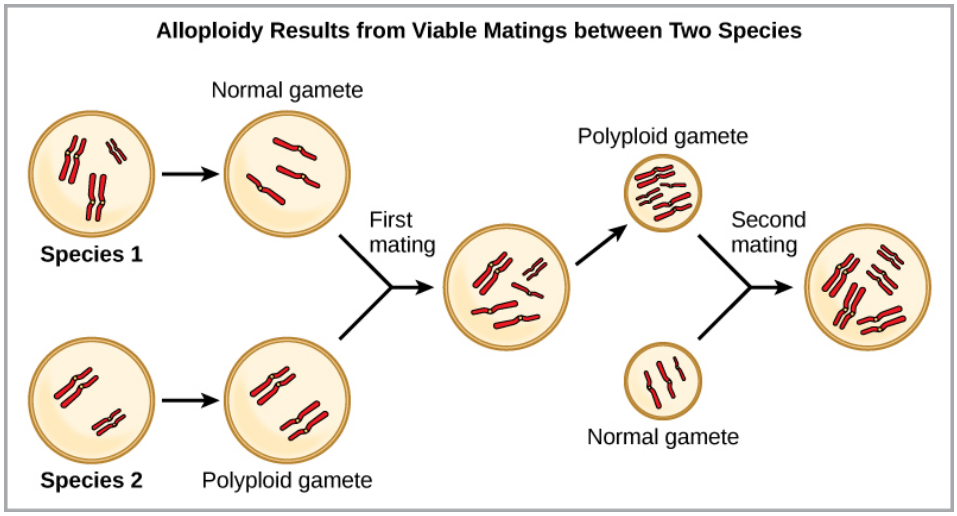


Figure 11.18 Alloploidy results when two species mate to produce viable offspring. In the example shown, a normal gamete from one species fuses with a polyploid gamete from another. Two matings are necessary to produce viable offspring.

The cultivated forms of wheat, cotton, and tobacco plants are all allopolyploids. Although polyploidy occurs occasionally in animals, most chromosomal abnormalities in animals are lethal; it takes place most commonly in plants. Scientists have discovered more than 1/2 of all plant species studied relate back to a species evolved through polyploidy.

Sympatric speciation may also take place in ways other than polyploidy. For example, imagine a species of fish that lived in a lake. As the population grew, competition for food also grew. Under pressure to find food, suppose that a group of these fish had the genetic flexibility to discover and feed off another resource that was unused by the other fish. What if this new food source was found at a different depth of the lake? Over time, those feeding on the second food source would interact more with each other than the other fish; therefore they would breed together as well. Offspring of these fish would likely behave as their parents and feed and live in the same area, keeping them separate from the original population. If this group of fish continued to remain separate from the first population, eventually sympatric speciation might occur as more genetic differences accumulated between them.

This scenario does play out in nature, as do others that lead to reproductive isolation. One such place is Lake Victoria in Africa, famous for its sympatric speciation of cichlid fish. Researchers have found hundreds of sympatric speciation events in these fish, which have not only happened in great number, but also over a short period of time. **Figure 11.19** shows this type of speciation among a cichlid fish population in Nicaragua. In this locale, two types of cichlids live in the same geographic location; however, they have come to have different morphologies that allow them to eat various food sources.

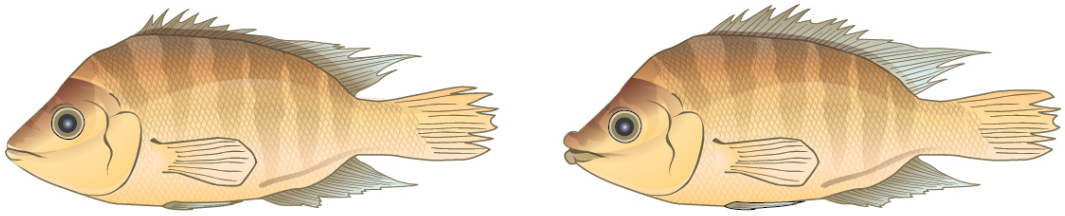


Figure 11.19 Cichlid fish from Lake Apoyeque, Nicaragua, show evidence of sympatric speciation. Lake Apoyeque, a crater lake, is 1800 years old, but genetic evidence indicates that the lake was populated only 100 years ago by a single population of cichlid fish. Nevertheless, two populations with distinct morphologies and diets now exist in the lake, and scientists believe these populations may be in an early stage of speciation.

Finally, a well-documented example of ongoing sympatric speciation occurred in the apple maggot fly, *Rhagoletis pomonella*, which arose as an isolated population sometime after the introduction of the apple into North America. The native population of flies fed on hawthorn species and is host-specific: it only infests hawthorn trees. Importantly, it also uses the trees as a location to meet for mating. It is hypothesized that either through mutation or a behavioral mistake, flies jumped hosts and met and mated in apple trees, subsequently laying their eggs in apple fruit. The offspring matured and kept their preference for the apple trees effectively dividing the original population into two new populations separated by host species, not by geography. The host jump took place in the nineteenth century, but there are now measureable differences between the two populations of fly. It seems likely that host specificity of parasites in general is a common cause of sympatric speciation.

11.5 | Common Misconceptions about Evolution

By the end of this section, you will be able to:

- Identify common misconceptions about evolution
- Identify common criticisms of evolution

Although the theory of evolution initially generated some controversy, by 20 years after the publication of *On the Origin of Species* it was almost universally accepted by biologists, particularly younger biologists. Nevertheless, the theory of evolution is a difficult concept and misconceptions about how it works abound. In addition, there are those that reject it as an explanation for the diversity of life.

This **website (http://openstaxcollege.org/l/misconception2)** addresses some of the main misconceptions associated with the theory of evolution.

Evolution Is Just a Theory

Critics of the theory of evolution dismiss its importance by purposefully confounding the everyday usage of the word "theory" with the way scientists use the word. In science, a "theory" is understood to be a concept that has been extensively tested and supported over time. We have a theory of the atom, a theory of gravity, and the theory of relativity, each of which describes what scientists understand to be facts about the world. In the same way, the theory of evolution describes facts about the living world. As such, a theory in science has survived significant efforts to discredit it by scientists, who are naturally skeptical. While theories can sometimes be overturned or revised, this does not lessen their weight but simply reflects the constantly evolving state of scientific knowledge. In contrast, a "theory" in common vernacular means a guess or suggested explanation for something. This meaning is more akin to the concept of a "hypothesis" used by scientists, which is a tentative explanation for something that is proposed to either be supported or disproved. When critics of evolution say evolution is "just a theory," they are implying that there is little evidence supporting it and that it is still in the process of being rigorously tested. This is a mischaracterization. If this were the case, geneticist Theodosius Dobzhansky would not have said that "nothing in biology makes sense, except in the light of evolution."[6]

Individuals Evolve

An individual is born with the genes it has—these do not change as the individual ages. Therefore, an individual cannot evolve or adapt through natural selection. Evolution is the change in genetic composition of a population over time, specifically over generations, resulting from differential reproduction of individuals with certain alleles. Individuals do change over their lifetime, but this is called development; it involves changes programmed by the set of genes the individual acquired at birth in coordination with the individual's environment. When thinking about the evolution of a characteristic, it is probably best to think about the change of the average value of the characteristic in the population over time. For example, when natural selection leads to bill-size change in medium ground finches in the Galápagos, this does not mean that individual bills on the finches are changing. If one measures the average bill size among all individuals in the population at one time, and then measures the average bill size in the population several years later after there has been a strong selective pressure, this average value may be different as a result of evolution. Although some individuals may survive from the first time to the second, those individuals will still have the same bill size. However, there may be enough new individuals with different bill sizes to change the average bill size.

Evolution Explains the Origin of Life

It is a common misunderstanding that evolution includes an explanation of life's origins. Conversely, some of the theory's critics complain that it cannot explain the origin of life. The theory does not try to explain the origin of life. The theory of evolution explains how populations change over time and how life diversifies—the origin of species. It does not shed light on the beginnings of life including the origins of the first cells, which is how life is defined. The mechanisms of the origin of life on Earth are a particularly difficult problem because it occurred a very long time ago, over a very long time, and presumably just occurred once. Importantly, biologists believe that the presence of life on Earth precludes the possibility that the events that led to life on Earth can be repeated because the intermediate stages would immediately become food for existing living things. The early stages of life included the formation of organic molecules such as carbohydrates, amino acids, or nucleotides. If these were formed from inorganic precursors today, they would simply be broken down by living things. The early stages of life also probably included more complex aggregations of molecules into enclosed structures with an internal environment, a boundary layer of some form, and the external environment. Such structures, if they were formed now, would be quickly consumed or broken down by living organisms.

6. Theodosius Dobzhansky. "Biology, Molecular and Organismic." *American Zoologist* 4, no. 4 (1964): 449.

However, once a mechanism of inheritance was in place in the form of a molecule like DNA or RNA, either within a cell or within a pre-cell, these entities would be subject to the principle of natural selection. More effective reproducers would increase in frequency at the expense of inefficient reproducers. So while evolution does not explain the origin of life, it may have something to say about some of the processes operating once pre-living entities acquired certain properties.

Organisms Evolve on Purpose

Statements such as "organisms evolve in response to a change in an environment," are quite common. There are two easy misunderstandings possible with such a statement. First of all, the statement must not be understood to mean that individual organisms evolve, as was discussed above. The statement is shorthand for "a population evolves in response to a changing environment." However, a second misunderstanding may arise by interpreting the statement to mean that the evolution is somehow intentional. A changed environment results in some individuals in the population, those with particular phenotypes, benefiting and, therefore, producing proportionately more offspring than other phenotypes. This results in change in the population if the characters are genetically determined.

It is also important to understand that the variation that natural selection works on is already in a population and does not arise in response to an environmental change. For example, applying antibiotics to a population of bacteria will, over time, select for a population of bacteria that are resistant to antibiotics. The resistance, which is caused by a gene, did not arise by mutation because of the application of the antibiotic. The gene for resistance was already present in the gene pool of the bacteria, likely at a low frequency. The antibiotic, which kills the bacterial cells without the resistance gene, strongly selects for individuals that are resistant, since these would be the only ones that survived and divided. Experiments have demonstrated that mutations for antibiotic resistance do not arise as a result of antibiotic application.

In a larger sense, evolution is also not goal directed. Species do not become "better" over time; they simply track their changing environment with adaptations that maximize their reproduction in a particular environment at a particular time. Evolution has no goal of making faster, bigger, more complex, or even smarter species. This kind of language is common in popular literature. Certain organisms, ourselves included, are described as the "pinnacle" of evolution, or "perfected" by evolution. What characteristics evolve in a species are a function of the variation present and the environment, both of which are constantly changing in a non-directional way. What trait is fit in one environment at one time may well be fatal at some point in the future. This holds equally well for a species of insect as it does the human species.

Evolution Is Controversial among Scientists

The theory of evolution was controversial when it was first proposed in 1859, yet within 20 years virtually every working biologist had accepted evolution as the explanation for the diversity of life. The rate of acceptance was extraordinarily rapid, partly because Darwin had amassed an impressive body of evidence. The early controversies involved both scientific arguments against the theory and the arguments of religious leaders. It was the arguments of the biologists that were resolved after a short time, while the arguments of religious leaders have persisted to this day.

The theory of evolution replaced the predominant theory at the time that species had all been specially created within relatively recent history. Despite the prevalence of this theory, it was becoming increasingly clear to naturalists during the nineteenth century that it could no longer explain many observations of geology and the living world. The persuasiveness of the theory of evolution to these naturalists lay in its ability to explain these phenomena, and it continues to hold extraordinary explanatory power to this day. Its continued rejection by some religious leaders results from its replacement of special creation, a tenet of their religious belief. These leaders cannot accept the replacement of special creation by a mechanistic process that excludes the actions of a deity as an explanation for the diversity of life including the origins of the human species. It should be noted, however, that most of the major denominations in the United States have statements supporting the acceptance of evidence for evolution as compatible with their theologies.

The nature of the arguments against evolution by religious leaders has evolved over time. One current argument is that the theory is still controversial among biologists. This claim is simply not true. The number of working scientists who reject the theory of evolution, or question its validity and say so, is small. A Pew Research poll in 2009 found that 97 percent of the 2500 scientists polled believe species evolve.[7] The support for the theory is reflected in signed statements from many scientific societies such as the American Association for the Advancement of Science, which includes working scientists as members. Many of the scientists that reject or question the theory of evolution are non-biologists, such as engineers, physicians, and chemists. There are no experimental results or research programs that contradict the theory. There are no papers published in peer-reviewed scientific journals that appear to refute the theory. The latter observation might be considered a consequence of suppression of dissent, but it must be remembered that scientists are skeptics and that there is a long history of published reports that challenged scientific orthodoxy in unpopular ways. Examples include the endosymbiotic theory of eukaryotic origins, the theory of group selection, the microbial cause of stomach ulcers, the

7. Pew Research Center for the People & the Press, *Public Praises Science; Scientists Fault Public, Media* (Washington, DC, 2009), 37.

asteroid-impact theory of the Cretaceous extinction, and the theory of plate tectonics. Research with evidence and ideas with scientific merit are considered by the scientific community. Research that does not meet these standards is rejected.

Other Theories Should Be Taught

A common argument from some religious leaders is that alternative theories to evolution should be taught in public schools. Critics of evolution use this strategy to create uncertainty about the validity of the theory without offering actual evidence. In fact, there are no viable alternative scientific theories to evolution. The last such theory, proposed by Lamarck in the nineteenth century, was replaced by the theory of natural selection. A single exception was a research program in the Soviet Union based on Lamarck's theory during the early twentieth century that set that country's agricultural research back decades. Special creation is not a viable alternative scientific theory because it is not a scientific theory, since it relies on an untestable explanation. Intelligent design, despite the claims of its proponents, is also not a scientific explanation. This is because intelligent design posits the existence of an unknown designer of living organisms and their systems. Whether the designer is unknown or supernatural, it is a cause that cannot be measured; therefore, it is not a scientific explanation. There are two reasons not to teach nonscientific theories. First, these explanations for the diversity of life lack scientific usefulness because they do not, and cannot, give rise to research programs that promote our understanding of the natural world. Experiments cannot test non-material explanations for natural phenomena. For this reason, teaching these explanations as science in public schools is not in the public interest. Second, in the United States, it is illegal to teach them as science because the U.S. Supreme Court and lower courts have ruled that the teaching of religious belief, such as special creation or intelligent design, violates the establishment clause of the First Amendment of the U.S. Constitution, which prohibits government sponsorship of a particular religion.

The theory of evolution and science in general is, by definition, silent on the existence or non-existence of the spiritual world. Science is only able to study and know the material world. Individual biologists have sometimes been vocal atheists, but it is equally true that there are many deeply religious biologists. Nothing in biology precludes the existence of a god, indeed biology as a science has nothing to say about it. The individual biologist is free to reconcile her or his personal and scientific knowledge as they see fit. The Voices for Evolution project (http://ncse.com/voices), developed through the National Center for Science Education, works to gather the diversity of perspectives on evolution to advocate it being taught in public schools.

KEY TERMS

adaptation a heritable trait or behavior in an organism that aids in its survival in its present environment

adaptive radiation a speciation when one species radiates out to form several other species

allopatric speciation a speciation that occurs via a geographic separation

analogous structure a structure that is similar because of evolution in response to similar selection pressures resulting in convergent evolution, not similar because of descent from a common ancestor

bottleneck effect the magnification of genetic drift as a result of natural events or catastrophes

convergent evolution an evolution that results in similar forms on different species

dispersal an allopatric speciation that occurs when a few members of a species move to a new geographical area

divergent evolution an evolution that results in different forms in two species with a common ancestor

founder effect a magnification of genetic drift in a small population that migrates away from a large parent population carrying with it an unrepresentative set of alleles

gene flow the flow of alleles in and out of a population due to the migration of individuals or gametes

gene pool all of the alleles carried by all of the individuals in the population

genetic drift the effect of chance on a population's gene pool

homologous structure a structure that is similar because of descent from a common ancestor

inheritance of acquired characteristics a phrase that describes the mechanism of evolution proposed by Lamarck in which traits acquired by individuals through use or disuse could be passed on to their offspring thus leading to evolutionary change in the population

macroevolution a broader scale of evolutionary changes seen over paleontological time

microevolution the changes in a population's genetic structure (i.e., allele frequency)

migration the movement of individuals of a population to a new location; in population genetics it refers to the movement of individuals and their alleles from one population to another, potentially changing allele frequencies in both the old and the new population

modern synthesis the overarching evolutionary paradigm that took shape by the 1940s and is generally accepted today

natural selection the greater relative survival and reproduction of individuals in a population that have favorable heritable traits, leading to evolutionary change

population genetics the study of how selective forces change the allele frequencies in a population over time

speciation a formation of a new species

sympatric speciation a speciation that occurs in the same geographic space

variation the variety of alleles in a population

vestigial structure a physical structure present in an organism but that has no apparent function and appears to be from a functional structure in a distant ancestor

vicariance an allopatric speciation that occurs when something in the environment separates organisms of the same species into separate groups

CHAPTER SUMMARY

11.1 Discovering How Populations Change

Evolution by natural selection arises from three conditions: individuals within a species vary, some of those variations are heritable, and organisms have more offspring than resources can support. The consequence is that individuals with relatively advantageous variations will be more likely to survive and have higher reproductive rates than those individuals with different traits. The advantageous traits will be passed on to offspring in greater proportion. Thus, the trait will have higher representation in the next and subsequent generations leading to genetic change in the population.

The modern synthesis of evolutionary theory grew out of the reconciliation of Darwin's, Wallace's, and Mendel's thoughts on evolution and heredity. Population genetics is a theoretical framework for describing evolutionary change in populations through the change in allele frequencies. Population genetics defines evolution as a change in allele frequency over generations. In the absence of evolutionary forces allele frequencies will not change in a population; this is known as Hardy-Weinberg equilibrium principle. However, in all populations, mutation, natural selection, genetic drift, and migration act to change allele frequencies.

11.2 Mechanisms of Evolution

There are four factors that can change the allele frequencies of a population. Natural selection works by selecting for alleles that confer beneficial traits or behaviors, while selecting against those for deleterious qualities. Mutations introduce new alleles into a population. Genetic drift stems from the chance occurrence that some individuals have more offspring than others and results in changes in allele frequencies that are random in direction. When individuals leave or join the population, allele frequencies can change as a result of gene flow.

11.3 Evidence of Evolution

The evidence for evolution is found at all levels of organization in living things and in the extinct species we know about through fossils. Fossils provide evidence for the evolutionary change through now extinct forms that led to modern species. For example, there is a rich fossil record that shows the evolutionary transitions from horse ancestors to modern horses that document intermediate forms and a gradual adaptation o changing ecosystems. The anatomy of species and the embryological development of that anatomy reveal common structures in divergent lineages that have been modified over time by evolution. The geographical distribution of living species reflects the origins of species in particular geographic locations and the history of continental movements. The structures of molecules, like anatomical structures, reflect the relationships of living species and match patterns of similarity expected from descent with modification.

11.4 Speciation

Speciation occurs along two main pathways: geographic separation (allopatric speciation) and through mechanisms that occur within a shared habitat (sympatric speciation). Both pathways force reproductive isolation between populations. Sympatric speciation can occur through errors in meiosis that form gametes with extra chromosomes, called polyploidy. Autopolyploidy occurs within a single species, whereas allopolyploidy occurs because of a mating between closely related species. Once the populations are isolated, evolutionary divergence can take place leading to the evolution of reproductive isolating traits that prevent interbreeding should the two populations come together again. The reduced viability of hybrid offspring after a period of isolation is expected to select for stronger inherent isolating mechanisms.

11.5 Common Misconceptions about Evolution

The theory of evolution is a difficult concept and misconceptions abound. The factual nature of evolution is often challenged by wrongly associating the scientific meaning of a theory with the vernacular meaning. Evolution is sometimes mistakenly interpreted to mean that individuals evolve, when in fact only populations can evolve as their gene frequencies change over time. Evolution is often assumed to explain the origin of life, which it does not speak to. It is often spoken in goal-directed terms by which organisms change through intention, and selection operates on mutations present in a population that have not arisen in response to a particular environmental stress. Evolution is often characterized as being controversial among scientists; however, it is accepted by the vast majority of working scientists. Critics of evolution often argue that alternative theories to evolution should be taught in public schools; however, there are no viable alternative scientific theories to evolution. The alternative religious beliefs should not be taught as science because it cannot be proven, and in the United States it is unconstitutional. Science is silent on the question of the existence of a god while scientists are able to reconcile religious belief and scientific knowledge.

ART CONNECTION QUESTIONS

1. **Figure 11.7** Do you think genetic drift would happen more quickly on an island or on the mainland?

REVIEW QUESTIONS

2. Which scientific concept did Charles Darwin and Alfred Wallace independently discover?

a. mutation
b. natural selection
c. overbreeding
d. sexual reproduction

3. Which of the following situations will lead to natural selection?

a. The seeds of two plants land near each other and one grows larger than the other.
b. Two types of fish eat the same kind of food, and one is better able to gather food than the other.
c. Male lions compete for the right to mate with females, with only one possible winner.
d. all of the above

4. What is the difference between micro- and macroevolution?

a. Microevolution describes the evolution of small organisms, such as insects, while macroevolution describes the evolution of large organisms, like people and elephants.
b. Microevolution describes the evolution of microscopic entities, such as molecules and proteins, while macroevolution describes the evolution of whole organisms.
c. Microevolution describes the evolution of populations, while macroevolution describes the emergence of new species over long periods of time.
d. Microevolution describes the evolution of organisms over their lifetimes, while macroevolution describes the evolution of organisms over multiple generations.

5. Population genetics is the study of _____.

a. how allele frequencies in a population change over time
b. populations of cells in an individual
c. the rate of population growth
d. how genes affect embryological development

6. Galápagos medium ground finches are found on Santa Cruz and San Cristóbal islands, which are separated by about 100 km of ocean. Occasionally, individuals from either island fly to the other island to stay. This can alter the allele frequencies of the population through which of the following mechanisms?

a. natural selection
b. genetic drift
c. gene flow
d. mutation

7. In which of the following pairs do both evolutionary processes introduce new genetic variation into a population?

a. natural selection and genetic drift
b. mutation and gene flow
c. natural selection and gene flow
d. gene flow and genetic drift

8. The wing of a bird and the arm of a human are examples of _____.

a. vestigial structures
b. molecular structures
c. homologous structures
d. analogous structures

9. The fact that DNA sequences are more similar in more closely related organisms is evidence of what?

a. optimal design in organisms
b. adaptation
c. mutation
d. descent with modification

10. Which situation would most likely lead to allopatric speciation?

a. A flood causes the formation of a new lake.
b. A storm causes several large trees to fall down.
c. A mutation causes a new trait to develop.
d. An injury causes an organism to seek out a new food source.

11. What is the main difference between dispersal and vicariance?

a. One leads to allopatric speciation, whereas the other leads to sympatric speciation.
b. One involves the movement of the organism, whereas the other involves a change in the environment.
c. One depends on a genetic mutation occurring, whereas the other does not.
d. One involves closely related organisms, whereas the other involves only individuals of the same species.

12. Which variable increases the likelihood of allopatric speciation taking place more quickly?

a. lower rate of mutation
b. longer distance between divided groups
c. increased instances of hybrid formation
d. equivalent numbers of individuals in each population

13. The word “theory” in theory of evolution is best replaced by _____.

a. fact
b. hypothesis

c. idea
d. alternate explanation

14. Why are alternative scientific theories to evolution not taught in public school?

a. more theories would confuse students
b. there are no viable scientific alternatives
c. it is against the law
d. alternative scientific theories are suppressed by the science establishment

CRITICAL THINKING QUESTIONS

15. If a person scatters a handful of plant seeds from one species in an area, how would natural selection work in this situation?

16. Explain the Hardy-Weinberg principle of equilibrium.

17. Describe natural selection and give an example of natural selection at work in a population.

18. Why do scientists consider vestigial structures evidence for evolution?

19. Why do island chains provide ideal conditions for adaptive radiation to occur?

20. Two species of fish had recently undergone sympatric speciation. The males of each species had a different coloring through which females could identify and choose a partner from her own species. After some time, pollution made the lake so cloudy it was hard for females to distinguish colors. What might take place in this situation?

21. How does the scientific meaning of “theory” differ from the common, everyday meaning of the word?

22. Explain why the statement that a monkey is more evolved than a mouse is incorrect.

12 | DIVERSITY OF LIFE

Figure 12.1 Although they look different, this bee and flower are distantly related. (credit: modification of work by John Beetham)

Chapter Outline

Introduction

This bee and *Echinacea* flower could not look more different, yet they are related, as are all living organisms on Earth. By following pathways of similarities and differences—both visible and genetic—scientists seek to map the history of evolution from single-celled organisms to the tremendous diversity of creatures that have crawled, germinated, floated, swam, flown, and walked on this planet.

12.1 | Organizing Life on Earth

By the end of this section, you will be able to:

- Discuss the need for a comprehensive classification system
- List the different levels of the taxonomic classification system
- Describe how systematics and taxonomy relate to phylogeny

All life on Earth evolved from a common ancestor. Biologists map how organisms are related by constructing phylogenetic trees. In other words, a “tree of life” can be constructed to illustrate when different organisms evolved and to show the relationships among different organisms, as shown in **Figure 12.2**. Notice that from a single point, the three domains of Archaea, Bacteria, and Eukarya diverge and then branch repeatedly. The small branch that plants and animals (including humans) occupy in this diagram shows how recently these groups had their origin compared with other groups.

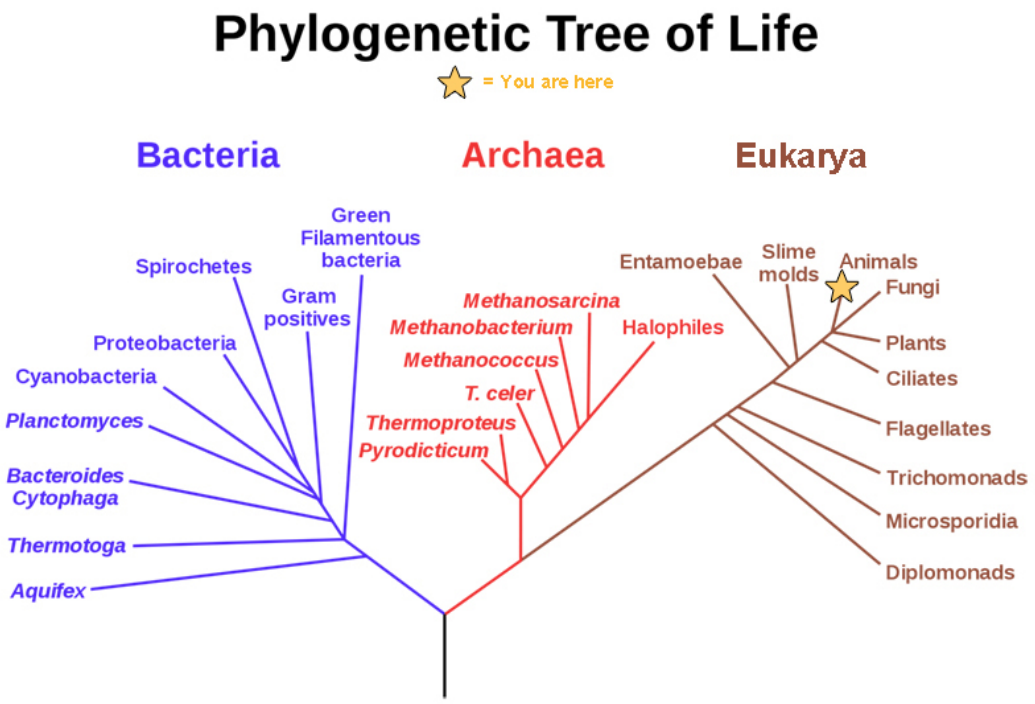


Figure 12.2 In the evolution of life on Earth, the three domains of life—Archaea, Bacteria, and Eukarya—branch from a single point. (credit: modification of work by Eric Gaba)

The phylogenetic tree in **Figure 12.2** illustrates the pathway of evolutionary history. The pathway can be traced from the origin of life to any individual species by navigating through the evolutionary branches between the two points. Also, by starting with a single species and tracing backward to any branch point, the organisms related to it by various degrees of closeness can be identified.

A **phylogeny** is the evolutionary history and the relationships among a species or group of species. The study of organisms with the purpose of deriving their relationships is called **systematics**.

Many disciplines within the study of biology contribute to understanding how past and present life evolved over time, and together they contribute to building, updating, and maintaining the "tree of life." Information gathered may include data collected from fossils, from studying morphology, from the structure of body parts, or from molecular structure, such as the sequence of amino acids in proteins or DNA nucleotides. By considering the trees generated by different sets of data scientists can put together the phylogeny of a species.

Scientists continue to discover new species of life on Earth as well as new character information, thus trees change as new data arrive.

The Levels of Classification

Taxonomy (which literally means "arrangement law") is the science of naming and grouping species to construct an internationally shared classification system. The taxonomic classification system (also called the Linnaean system after its inventor, Carl Linnaeus, a Swedish naturalist) uses a hierarchical model. A hierarchical system has levels and each group at one of the levels includes groups at the next lowest level, so that at the lowest level each member belongs to a series of nested groups. An analogy is the nested series of directories on the main disk drive of a computer. For example, in the most inclusive grouping, scientists divide organisms into three **domains**: Bacteria, Archaea, and Eukarya. Within each domain is a second level called a **kingdom**. Each domain contains several kingdoms. Within kingdoms, the subsequent categories of increasing specificity are: **phylum**, **class**, **order**, **family**, **genus**, and **species**.

As an example, the classification levels for the domestic dog are shown in **Figure 12.3**. The group at each level is called a **taxon** (plural: taxa). In other words, for the dog, Carnivora is the taxon at the order level, Canidae is the taxon at the family level, and so forth. Organisms also have a common name that people typically use, such as domestic dog, or wolf. Each taxon name is capitalized except for species, and the genus and species names are italicized. Scientists refer to an organism by its genus and species names together, commonly called a scientific name, or Latin name. This two-name system is called **binomial nomenclature**. The scientific name of the wolf is therefore *Canis lupus*. Recent study of the DNA of domestic

dogs and wolves suggest that the domestic dog is a subspecies of the wolf, not its own species, thus it is given an extra name to indicate its subspecies status, *Canis lupus familiaris*.

Figure 12.3 also shows how taxonomic levels move toward specificity. Notice how within the domain we find the dog grouped with the widest diversity of organisms. These include plants and other organisms not pictured, such as fungi and protists. At each sublevel, the organisms become more similar because they are more closely related. Before Darwin's theory of evolution was developed, naturalists sometimes classified organisms using arbitrary similarities, but since the theory of evolution was proposed in the 19th century, biologists work to make the classification system reflect evolutionary relationships. This means that all of the members of a taxon should have a common ancestor and be more closely related to each other than to members of other taxa.

Recent genetic analysis and other advancements have found that some earlier taxonomic classifications do not reflect actual evolutionary relationships, and therefore, changes and updates must be made as new discoveries take place. One dramatic and recent example was the breaking apart of prokaryotic species, which until the 1970s were all classified as bacteria. Their division into Archaea and Bacteria came about after the recognition that their large genetic differences warranted their separation into two of three fundamental branches of life.

art CONNECTION

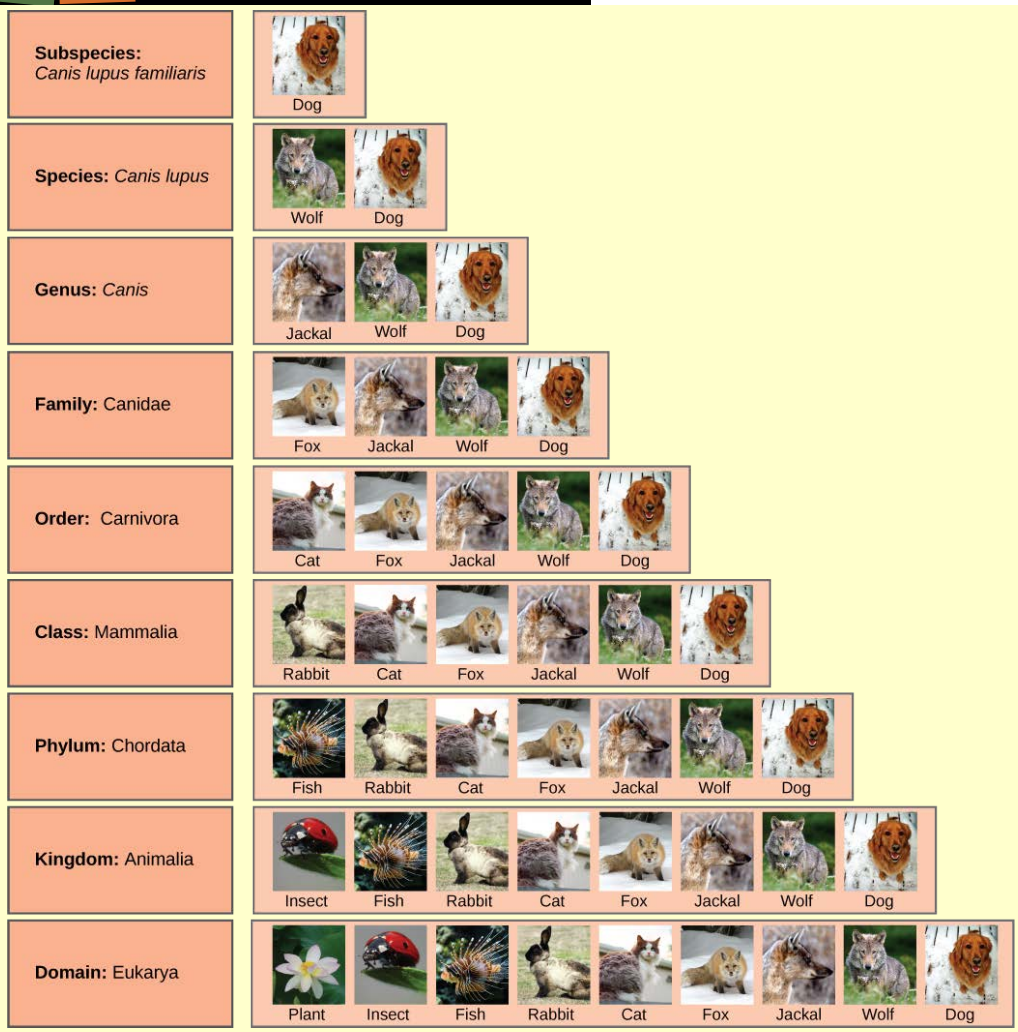


Figure 12.3 At each sublevel in the taxonomic classification system, organisms become more similar. Dogs and wolves are the same species because they can breed and produce viable offspring, but they are different enough to be classified as different subspecies. (credit “plant”: modification of work by "berduchwal"/Flickr; credit “insect”: modification of work by Jon Sullivan; credit “fish”: modification of work by Christian Mehlführer; credit “rabbit”: modification of work by Aidan Wojtas; credit “cat”: modification of work by Jonathan Lidbeck; credit “fox”: modification of work by Kevin Bacher, NPS; credit “jackal”: modification of work by Thomas A. Hermann, NBII, USGS; credit “wolf” modification of work by Robert Dewar; credit “dog”: modification of work by "digital_image_fan"/Flickr)

In what levels are cats and dogs considered to be part of the same group?

Visit **this PBS site (http://openstaxcollege.org/l/classify_life2)** to learn more about taxonomy. Under Classifying Life, click Launch Interactive.

Classification and Phylogeny

Scientists use a tool called a phylogenetic tree to show the evolutionary pathways and relationships between organisms. A **phylogenetic tree** is a diagram used to reflect evolutionary relationships among organisms or groups of organisms. The hierarchical classification of groups nested within more inclusive groups is reflected in diagrams. Scientists consider phylogenetic trees to be a hypothesis of the evolutionary past because one cannot go back through time to confirm the proposed relationships.

Unlike with a taxonomic classification, a phylogenetic tree can be read like a map of evolutionary history, as shown in **Figure 12.4**. Shared characteristics are used to construct phylogenetic trees. The point where a split occurs in a tree, called a **branch point**, represents where a single lineage evolved into distinct new ones. Many phylogenetic trees have a single branch point at the base representing a common ancestor of all the branches in the tree. Scientists call such trees **rooted**, which means there is a single ancestral taxon at the base of a phylogenetic tree to which all organisms represented in the diagram descend from. When two lineages stem from the same branch point, they are called **sister taxa**, for example the two species of orangutans. A branch point with more than two groups illustrates a situation for which scientists have not definitively determined relationships. An example is illustrated by the three branches leading to the gorilla subspecies; their exact relationships are not yet understood. It is important to note that sister taxa share an ancestor, which does not mean that one taxon evolved from the other. The branch point, or split, represents a common ancestor that existed in the past, but that no longer exists. Humans did not evolve from chimpanzees (nor did chimpanzees evolve from humans) although they are our closest living relatives. Both humans and chimpanzees evolved from a common ancestor that lived, scientists believe, six million years ago and looked different from both modern chimpanzees and modern humans.

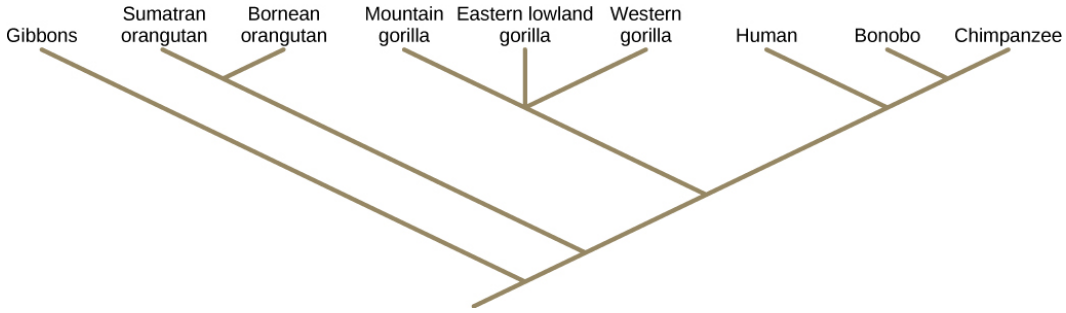


Figure 12.4 A phylogenetic tree is rooted and shows how different organisms, in this case the species and subspecies of living apes, evolved from a common ancestor.

The branch points and the branches in phylogenetic tree structure also imply evolutionary change. Sometimes the significant character changes are identified on a branch or branch point. For example, in **Figure 12.5**, the branch point that gives rise to the mammal and reptile lineage from the frog lineage shows the origin of the amniotic egg character. Also the branch point that gives rise to organisms with legs is indicated at the common ancestor of mammals, reptiles, amphibians, and jawed fishes.

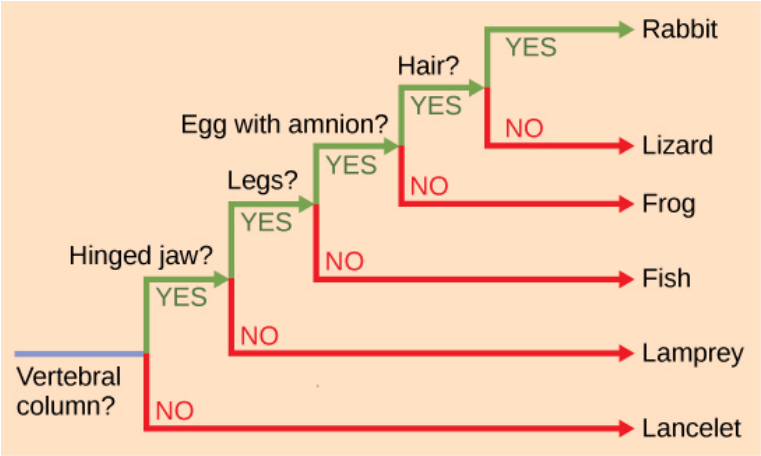


Figure 12.5 This phylogenetic tree is rooted by an organism that lacked a vertebral column. At each branch point, organisms with different characters are placed in different groups.

CONCEPT in ACTION

openstax COLLEGE

This **interactive exercise (http://openstaxcollege.org/l/tree_of_life3)** allows you to explore the evolutionary relationships among species.

Limitations of Phylogenetic Trees

It is easy to assume that more closely related organisms look more alike, and while this is often the case, it is not always true. If two closely related lineages evolved under significantly different surroundings or after the evolution of a major new adaptation, they may look quite different from each other, even more so than other groups that are not as closely related. For example, the phylogenetic tree in **Figure 12.5** shows that lizards and rabbits both have amniotic eggs, whereas salamanders (within the frog lineage) do not; yet on the surface, lizards and salamanders appear more similar than the lizards and rabbits.

Another aspect of phylogenetic trees is that, unless otherwise indicated, the branches do not show length of time, they show only the order in time of evolutionary events. In other words, a long branch does not necessarily mean more time passed, nor does a short branch mean less time passed— unless specified on the diagram. For example, in **Figure 12.5**, the tree does not indicate how much time passed between the evolution of amniotic eggs and hair. What the tree does show is the order in which things took place. Again using **Figure 12.5**, the tree shows that the oldest trait is the vertebral column, followed by hinged jaws, and so forth. Remember that any phylogenetic tree is a part of the greater whole, and similar to a real tree, it does not grow in only one direction after a new branch develops. So, for the organisms in **Figure 12.5**, just because a vertebral column evolved does not mean that invertebrate evolution ceased, it only means that a new branch formed. Also, groups that are not closely related, but evolve under similar conditions, may appear more similar to each other than to a close relative.

12.2 | Determining Evolutionary Relationships

By the end of this section, you will be able to:

- Compare homologous and analogous traits
- Discuss the purpose of cladistics

Scientists collect information that allows them to make evolutionary connections between organisms. Similar to detective work, scientists must use evidence to uncover the facts. In the case of phylogeny, evolutionary investigations focus on two types of evidence: morphologic (form and function) and genetic.

Two Measures of Similarity

Organisms that share similar physical features *and* genetic sequences tend to be more closely related than those that do not. Features that overlap both morphologically and genetically are referred to as homologous structures; the similarities stem from common evolutionary paths. For example, as shown in **Figure 12.6**, the bones in the wings of bats and birds, the arms of humans, and the foreleg of a horse are homologous structures. Notice the structure is not simply a single bone, but rather a grouping of several bones arranged in a similar way in each organism even though the elements of the structure may have changed shape and size.

Figure 12.6 Bat and bird wings, the foreleg of a horse, the flipper of a whale, and the arm of a human are homologous structures, indicating that bats, birds, horses, whales, and humans share a common evolutionary past. (credit a photo: modification of work by Steve Hillebrand, USFWS; credit b photo: modification of work by U.S. BLM; credit c photo: modification of work by Virendra Kankariya; credit d photo: modification of work by Russian Gov./Wikimedia Commons)

Misleading Appearances

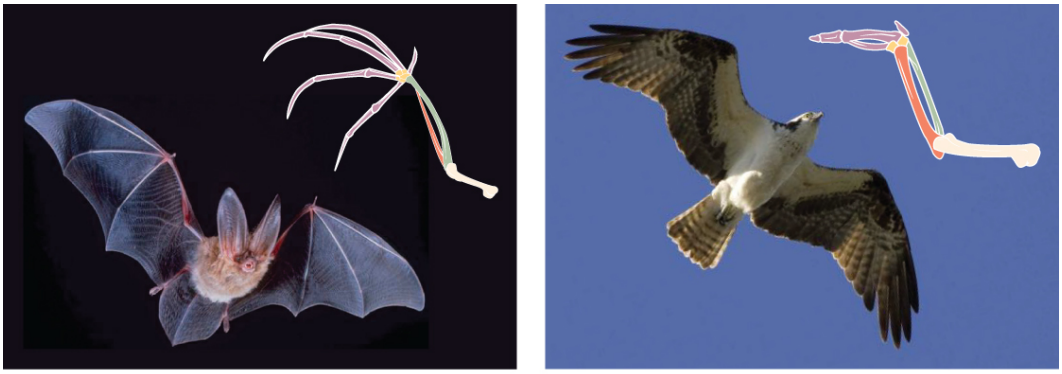
Some organisms may be very closely related, even though a minor genetic change caused a major morphological difference to make them look quite different. For example, chimpanzees and humans, the skulls of which are shown in **Figure 12.7** are very similar genetically, sharing 99 percent [1] of their genes. However, chimpanzees and humans show considerable anatomical differences, including the degree to which the jaw protrudes in the adult and the relative lengths of our arms and legs.

1. Gibbons, A. (2012, June 13). *Science Now*. Retrieved from http://news.sciencemag.org/sciencenow/2012/06/bonobo-genome-sequenced.html

Figure 12.7 (a) The chimpanzee jaw protrudes to a much greater degree than (b) the human jaw. (credit a: modification of work by "Pastorius"/Wikimedia Commons)

However, unrelated organisms may be distantly related yet appear very much alike, usually because common adaptations to similar environmental conditions evolved in both. An example is the streamlined body shapes, the shapes of fins and appendages, and the shape of the tails in fishes and whales, which are mammals. These structures bear superficial similarity because they are adaptations to moving and maneuvering in the same environment—water. When a characteristic that is similar occurs by adaptive convergence (convergent evolution), and not because of a close evolutionary relationship, it is called an **analogous structure**. In another example, insects use wings to fly like bats and birds. We call them both wings because they perform the same function and have a superficially similar form, but the embryonic origin of the two wings is completely different. The difference in the development, or embryogenesis, of the wings in each case is a signal that insects and bats or birds do not share a common ancestor that had a wing. The wing structures, shown in **Figure 12.8** evolved independently in the two lineages.

Similar traits can be either homologous or analogous. Homologous traits share an evolutionary path that led to the development of that trait, and analogous traits do not. Scientists must determine which type of similarity a feature exhibits to decipher the phylogeny of the organisms being studied.



(a) Bat wing (b) Bird wing

(c) Insect wing

Figure 12.8 The wing of a honey bee is similar in shape to a bird wing and a bat wing and serves the same function (flight). The bird and bat wings are homologous structures. However, the honey bee wing has a different structure (it is made of a chitinous exoskeleton, not a boney endoskeleton) and embryonic origin. The bee and bird or bat wing types illustrate an analogy—similar structures that do not share an evolutionary history. (credit a photo: modification of work by U.S. BLM; credit b: modification of work by Steve Hillebrand, USFWS; credit c: modification of work by Jon Sullivan)

CONCEPT in ACTION

This **website (http://openstaxcollege.org/l/relationships2)** has several examples to show how appearances can be misleading in understanding the phylogenetic relationships of organisms.

Molecular Comparisons

With the advancement of DNA technology, the area of **molecular systematics**, which describes the use of information on the molecular level including DNA sequencing, has blossomed. New analysis of molecular characters not only confirms many earlier classifications, but also uncovers previously made errors. Molecular characters can include differences in the amino-acid sequence of a protein, differences in the individual nucleotide sequence of a gene, or differences in the arrangements of genes. Phylogenies based on molecular characters assume that the more similar the sequences are in two organisms, the more closely related they are. Different genes change evolutionarily at different rates and this affects the level at which they are useful at identifying relationships. Rapidly evolving sequences are useful for determining the relationships among closely related species. More slowly evolving sequences are useful for determining the relationships between distantly related species. To determine the relationships between very different species such as Eukarya and Archaea, the genes used must be very ancient, slowly evolving genes that are present in both groups, such as the genes for

ribosomal RNA. Comparing phylogenetic trees using different sequences and finding them similar helps to build confidence in the inferred relationships.

Sometimes two segments of DNA in distantly related organisms randomly share a high percentage of bases in the same locations, causing these organisms to appear closely related when they are not. For example, the fruit fly shares 60 percent of its DNA with humans.[2] In this situation, computer-based statistical algorithms have been developed to help identify the actual relationships, and ultimately, the coupled use of both morphologic and molecular information is more effective in determining phylogeny.

evolution IN ACTION

Why Does Phylogeny Matter?

In addition to enhancing our understanding of the evolutionary history of species, our own included, phylogenetic analysis has numerous practical applications. Two of those applications include understanding the evolution and transmission of disease and making decisions about conservation efforts. A 2010 study[3] of MRSA (methicillin-resistant *Staphylococcus aureus*), an antibiotic resistant pathogenic bacterium, traced the origin and spread of the strain throughout the past 40 years. The study uncovered the timing and patterns in which the resistant strain moved from its point of origin in Europe to centers of infection and evolution in South America, Asia, North America, and Australasia. The study suggested that introductions of the bacteria to new populations occurred very few times, perhaps only once, and then spread from that limited number of individuals. This is in contrast to the possibility that many individuals had carried the bacteria from one place to another. This result suggests that public health officials should concentrate on quickly identifying the contacts of individuals infected with a new strain of bacteria to control its spread.

A second area of usefulness for phylogenetic analysis is in conservation. Biologists have argued that it is important to protect species throughout a phylogenetic tree rather than just those from one branch of the tree. Doing this will preserve more of the variation produced by evolution. For example, conservation efforts should focus on a single species without sister species rather than another species that has a cluster of close sister species that recently evolved. If the single evolutionarily distinct species goes extinct a disproportionate amount of variation from the tree will be lost compared to one species in the cluster of closely related species. A study published in 2007[4] made recommendations for conservation of mammal species worldwide based on how evolutionarily distinct and at risk of extinction they are. The study found that their recommendations differed from priorities based on simply the level of extinction threat to the species. The study recommended protecting some threatened and valued large mammals such as the orangutans, the giant and lesser pandas, and the African and Asian elephants. But they also found that some much lesser known species should be protected based on how evolutionary distinct they are. These include a number of rodents, bats, shrews and hedgehogs. In addition there are some critically endangered species that did not rate as very important in evolutionary distinctiveness including species of deer mice and gerbils. While many criteria affect conservation decisions, preserving phylogenetic diversity provides an objective way to protect the full range of diversity generated by evolution.

Building Phylogenetic Trees

How do scientists construct phylogenetic trees? Presently, the most accepted method for constructing phylogenetic trees is a method called **cladistics**. This method sorts organisms into **clades**, groups of organisms that are most closely related to each other and the ancestor from which they descended. For example, in **Figure 12.9**, all of the organisms in the shaded region evolved from a single ancestor that had amniotic eggs. Consequently, all of these organisms also have amniotic eggs and make a single clade, also called a **monophyletic group**. Clades must include the ancestral species and all of the descendants from a branch point.

2. *Background on comparative genomic analysis*. (2002, December). Retrieved from http://www.genome.gov/10005835
3. Harris, S.R. et al. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469–474.
4. Isaac NJ, Turvey ST, Collen B, Waterman C, Baillie JE (2007) Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. PLoS ONE 2(3): e296. doi:10.1371/journal.pone.0000296

art CONNECTION

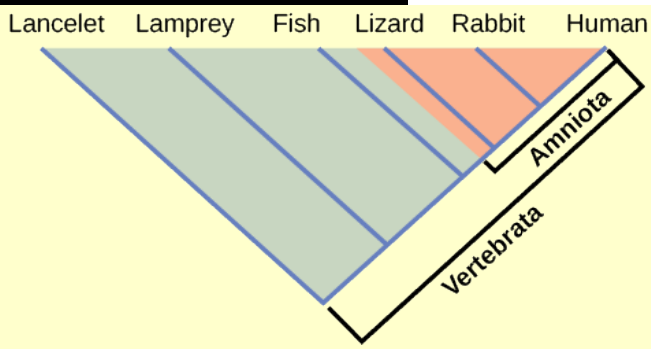


Figure 12.9 Lizards, rabbits, and humans all descend from a common ancestor in which the amniotic egg evolved. Thus, lizards, rabbits, and humans all belong to the clade Amniota. Vertebrata is a larger clade that also includes fish and lamprey.

Which animals in this figure belong to a clade that includes animals with hair? Which evolved first: hair or the amniotic egg?

Clades can vary in size depending on which branch point is being referenced. The important factor is that all of the organisms in the clade or monophyletic group stem from a single point on the tree. This can be remembered because monophyletic breaks down into "mono," meaning one, and "phyletic," meaning evolutionary relationship.

Shared Characteristics

Cladistics rests on three assumptions. The first is that living things are related by descent from a common ancestor, which is a general assumption of evolution. The second is that speciation occurs by splits of one species into two, never more than two at a time, and essentially at one point in time. This is somewhat controversial, but is acceptable to most biologists as a simplification. The third assumption is that traits change enough over time to be considered to be in a different state .It is also assumed that one can identify the actual direction of change for a state. In other words, we assume that an amniotic egg is a later character state than non-amniotic eggs. This is called the polarity of the character change. We know this by reference to a group outside the clade: for example, insects have non-amniotic eggs; therefore, this is the older or ancestral character state. Cladistics compares ingroups and outgroups. An ingroup (lizard, rabbit and human in our example) is the group of taxa being analyzed. An outgroup (lancelet, lamprey and fish in our example) is a species or group of species that diverged before the lineage containing the group(s) of interest. By comparing ingroup members to each other and to the outgroup members, we can determine which characteristics are evolutionary modifications determining the branch points of the ingroup's phylogeny.

If a characteristic is found in all of the members of a group, it is a **shared ancestral character** because there has been no change in the trait during the descent of each of the members of the clade. Although these traits appear interesting because they unify the clade, in cladistics they are considered not helpful when we are trying to determine the relationships of the members of the clade because every member is the same. In contrast, consider the amniotic egg characteristic of **Figure 12.9**. Only some of the organisms have this trait, and to those that do, it is called a **shared derived character** because this trait changed at some point during descent. This character does tell us about the relationships among the members of the clade; it tells us that lizards, rabbits, and humans group more closely together than any of these organisms do with fish, lampreys, and lancelets.

A sometimes confusing aspect of "ancestral" and "derived" characters is that these terms are relative. The same trait could be either ancestral or derived depending on the diagram being used and the organisms being compared. Scientists find these terms useful when distinguishing between clades during the building of phylogenetic trees, but it is important to remember that their meaning depends on context.

Choosing the Right Relationships

Constructing a phylogenetic tree, or cladogram, from the character data is a monumental task that is usually left up to a computer. The computer draws a tree such that all of the clades share the same list of derived characters. But there are other decisions to be made, for example, what if a species presence in a clade is supported by all of the shared derived characters for that clade except one? One conclusion is that the trait evolved in the ancestor, but then changed back in that one species. Also a character state that appears in two clades must be assumed to have evolved independently in those clades. These

inconsistencies are common in trees drawn from character data and complicate the decision-making process about which tree most closely represents the real relationships among the taxa.

To aid in the tremendous task of choosing the best tree, scientists often use a concept called **maximum parsimony**, which means that events occurred in the simplest, most obvious way. This means that the "best" tree is the one with the fewest number of character reversals, the fewest number of independent character changes, and the fewest number of character changes throughout the tree. Computer programs search through all of the possible trees to find the small number of trees with the simplest evolutionary pathways. Starting with all of the homologous traits in a group of organisms, scientists can determine the order of evolutionary events of which those traits occurred that is the most obvious and simple.

Practice Parsimony: Go to **this website (http://openstaxcollege.org/l/parsimony2)** to learn how maximum parsimony is used to create phylogenetic trees (be sure to continue to the second page).

These tools and concepts are only a few of the strategies scientists use to tackle the task of revealing the evolutionary history of life on Earth. Recently, newer technologies have uncovered surprising discoveries with unexpected relationships, such as the fact that people seem to be more closely related to fungi than fungi are to plants. Sound unbelievable? As the information about DNA sequences grows, scientists will become closer to mapping the evolutionary history of all life on Earth.

KEY TERMS

analogous structure a character found in two taxa that looks similar because of convergent evolution, not because of descent from a common ancestor

binomial nomenclature a system of two-part scientific names for an organism, which includes genus and species names

branch point a point on a phylogenetic tree where a single lineage splits to distinct new ones

clade a group of taxa with the same set of shared derived characters, including an ancestral species and all its descendants

cladistics a method used to organize homologous traits to describe phylogenies using common descendent as the primary criterion used to classify organisms

class the category in the taxonomic classification system that falls within phylum and includes orders

domain the highest level category in the classification system and that includes all taxonomic classifications below it; it is the most inclusive taxon

family the category in the taxonomic classification system that falls within order and includes genera

genus the category in the taxonomic classification system that falls within family and includes species; the first part of the scientific name

kingdom the category in the taxonomic classification system that falls within domain and includes phyla

maximum parsimony applying the simplest, most obvious way with the least number of steps

molecular systematics the methods of using molecular evidence to identify phylogenetic relationships

monophyletic group (also, clade) organisms that share a single ancestor

order the category in the taxonomic classification system that falls within class and includes families

phylogenetic tree diagram used to reflect the evolutionary relationships between organisms or groups of organisms

phylogeny evolutionary history and relationship of an organism or group of organisms

phylum the category in the taxonomic classification system that falls within kingdom and includes classes

rooted describing a phylogenetic tree with a single ancestral lineage to which all organisms represented in the diagram relate

shared ancestral character a character on a phylogenetic branch that is shared by a particular clade

shared derived character a character on a phylogenetic tree that is shared only by a certain clade of organisms

sister taxa two lineages that diverged from the same branch point

species the most specific category of classification

systematics the science of determining the evolutionary relationships of organisms

taxon a single level in the taxonomic classification system

taxonomy the science of classifying organisms

CHAPTER SUMMARY

12.1 Organizing Life on Earth

Scientists continually obtain new information that helps to understand the evolutionary history of life on Earth. Each group of organisms went through its own evolutionary journey, called its phylogeny. Each organism shares relatedness with

others, and based on morphologic and genetic evidence scientists attempt to map the evolutionary pathways of all life on Earth. Historically, organisms were organized into a taxonomic classification system. However, today many scientists build phylogenetic trees to illustrate evolutionary relationships and the taxonomic classification system is expected to reflect evolutionary relationships.

12.2 Determining Evolutionary Relationships

To build phylogenetic trees, scientists must collect character information that allows them to make evolutionary connections between organisms. Using morphologic and molecular data, scientists work to identify homologous characteristics and genes. Similarities between organisms can stem either from shared evolutionary history (homologies) or from separate evolutionary paths (analogies). After homologous information is identified, scientists use cladistics to organize these events as a means to determine an evolutionary timeline. Scientists apply the concept of maximum parsimony, which states that the likeliest order of events is probably the simplest shortest path. For evolutionary events, this would be the path with the least number of major divergences that correlate with the evidence.

ART CONNECTION QUESTIONS

1. **Figure 12.3** In what levels are cats and dogs considered to be part of the same group?

2. **Figure 12.8** Which animals in this figure belong to a clade that includes animals with hair? Which evolved first: hair or the amniotic egg?

REVIEW QUESTIONS

3. What is a phylogeny a description of?
- a. mutations
- b. DNA
- c. evolutionary history
- d. organisms on Earth

4. What do scientists in the field of systematics accomplish?
- a. discover new fossil sites
- b. organize and classify organisms
- c. name new species
- d. communicate between field biologists

5. Which statement about the taxonomic classification system is correct?
- a. There are more domains than kingdoms.
- b. Kingdoms are the top category of classification.
- c. A phylum may be represented in more than one kingdom.
- d. Species are the most specific category of classification.

6. Which best describes the relationship between chimpanzees and humans?
- a. chimpanzees evolved from humans
- b. humans evolved from chimpanzees
- c. chimpanzees and humans evolved from a common ancestor
- d. chimpanzees and humans belong to the same species

7. Which best describes a branch point in a phylogenetic tree?
- a. a hypothesis
- b. new lineage
- c. hybridization
- d. a mating

8. Which statement about analogies is correct?
- a. They occur only as errors.
- b. They are synonymous with homologous traits.
- c. They are derived by response to similar environmental pressures.
- d. They are a form of mutation.

9. What kind of trait is important to cladistics?
- a. shared derived traits
- b. shared ancestral traits
- c. analogous traits
- d. parsimonious traits

10. What is true about organisms that are a part of the same clade?
- a. They all share the same basic characteristics.
- b. They evolved from a shared ancestor.
- c. They all are on the same tree.
- d. They have identical phylogenies.

11. Which assumption of cladistics is stated incorrectly?
- a. Living things are related by descent from a common ancestor.
- b. Speciation can produce one, two, or three new species.
- c. Traits change from one state to another.
- d. The polarity of a character state change can be determined.

12. A monophyletic group is a _____.
- a. phylogenetic tree
- b. shared derived trait
- c. character state
- d. clade

CRITICAL THINKING QUESTIONS

13. How does a phylogenetic tree indicate major evolutionary events within a lineage?

14. List the different levels of the taxonomic classification system.

15. Dolphins and fish have similar body shapes. Is this feature more likely a homologous or analogous trait?

16. Describe maximum parsimony.

17. How does a biologist determine the polarity of a character change?